

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:22:25 ; Search time 97.3985 Seconds  
(without alignments)  
2272.479 Million cell updates/sec

Title: US-10-658-989A-2  
Perfect score: 3488  
Sequence: 1 GPGEPGPTGLPFPBGRGG.....GEGVFGDLGAPGSPGAGG 617

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_23Sep04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3488	100.0	617	8 ADM48391	Adm48391 Human rec
2	3488	100.0	617	8 ADQ26217	Adq26217 Human gel
3	3482	99.8	821	8 ADM48392	Adm48392 Human rec
4	2237.5	64.1	1057	3 AAY84544	Aay84544 A human c
5	2237.5	64.1	1058	3 AAY84403	Aay84403 Amino aci
6	2237.5	64.1	1107	3 AAY84540	Aay84540 Amino aci
7	2237.5	64.1	1171	3 AAY84538	Aay84538 A chimeri
8	2237.5	64.1	1464	2 AAW68485	Aaw68485 Human rec
9	2237.5	64.1	1464	7 ADD45059	Add45059 Human Pro
10	2237.5	64.1	1464	7 ADD45055	Add45055 Human Pro
11	2237.5	64.1	1464	7 ADD45051	Add45051 Human Pro
12	2234	64.0	1449	4 AAE02535	Aae02535 Porcine a
13	2233.5	64.0	1161	7 ADE87050	Ade87050 Human pan
14	2233.5	64.0	1461	5 AEG93947	Aeg93947 Human pol
15	2233.5	64.0	1464	4 AAUI1436	Aaui1436 Human nov
16	2233.5	64.0	1536	7 ADE87051	Ade87051 Human pan
17	2231.5	64.0	1057	3 AAY84541	Aay84541 Amino aci
18	2229.5	63.9	1388	3 AAY84539	Aay84539 Amino aci
19	2228.5	63.9	1341	2 AAR71701	Aar71701 Collagen
20	2228.5	63.9	1341	3 AAY96122	Aay96122 Collagen
21	2228.5	63.9	1341	5 AAE18475	Aae18475 Human col
22	2228.5	63.9	1341	5 ABB80733	Abb80733 Collagen
23	2228.5	63.9	1341	5 ABB03625	Abb03625 Amino aci
24	2228.5	63.9	1341	7 ADF13075	Adf13075 Human col
25	2227.5	63.9	1464	4 AAB82454	Aab82454 Human pro

26	2227.5	63.9	1464	5 ABB90764	Abb90764 Human tum
27	2227.5	63.9	1464	5 ABP68610	Abp68610 Human pan
28	2227.5	63.9	1464	6 ABUS4471	Abus4471 Human tum
29	2227.5	63.9	1464	6 ABR47417	Abx47417 Breast ca
30	2227.5	63.9	1464	6 ABR92064	Abx92064 Human cer
31	2227.5	63.9	1464	7 ADD14142	Ad14142 Human src
32	2227.5	63.9	1464	7 ADF65246	Adf65246 Human alp
33	2227.5	63.9	1464	8 ADQ19470	Adq19470 Human sof
34	2225.5	63.8	1463	4 AAE02532	Aae02532 Bovine al
35	2224.5	63.8	1464	7 ADE87048	Ade87048 Human pan
36	2222.5	63.7	1107	2 AAR89472	Aar89472 Collagen/
37	2222.5	63.7	1169	2 AAR89469	Aar89469 Amino aci
38	2222.5	63.7	1169	3 AAY84537	Aay84537 Amino aci
39	2222.5	63.7	1171	2 AAR89470	Aar89470 Collagen/
40	2222.5	63.7	1388	2 AAR89471	Aar89471 Collagen/
41	2222	63.7	1211	7 ADE87057	Ade87057 Human pan
42	2222	63.7	1226	7 ADE87062	Ade87062 Human pan
43	2222	63.7	1411	3 AAY56800	Aay56800 Human pre
44	2215.5	63.5	1453	7 ADD45053	Add45053 Rat Prote
45	2215.5	63.5	1453	7 ADD45057	Add45057 Rat Prote

ALIGNMENTS

RESULT 1  
ADM48391  
ID ADM48391 standard; protein; 617 AA.  
XX  
AC ADM48391;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE Human recombinant gelatin-like polypeptide Hu-3.  
XX  
KW Plasma substitute; Gelatin-like protein; plasma expander; human.  
XX  
OS Homo sapiens.  
XX  
PN EP1398324-A1.  
XX  
PD 17-MAR-2004.  
XX  
PF 11-SEP-2002; 2002EP-00078745.  
XX  
PR 11-SEP-2002; 2002EP-00078745.  
XX  
PA (FUJF ) FUJI PHOTO FILM BV.  
XX  
PI Bouwstra JB, Toda Y;  
DR WPI; 2004-229415/22.  
XX  
PT Composition useful as substitute for plasma, comprises solution of saline  
PT and recombinant gelatin-like protein having colloid osmotic function.  
XX  
Example 1; SEQ ID NO 2; 3lpp; English.  
The present sequence is the protein sequence of recombinant gelatin-like protein Hu-3. This is a trimer of human recombinant gelatin-like protein Hu-1 ADM48390. A claimed composition suitable as a substitute for plasma comprises a solution of saline and a protein having a colloid osmotic function. The protein is a recombinant gelatin-like protein with a molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The recombinant gelatin-like protein, or a dimer, trimer or tetramer of the protein, is useful as a plasma expander that has a lower clearance rate from blood circulation, provides better and predictable regulation of clearance rate and which is less susceptible to proteolytic degradation than presently used gelatin derivatives. Recombinant gelatin-like proteins that are in essence free of hydroxyproline do not give rise to an immunological reaction with blood samples containing IGE antibodies. The gelatin-like proteins can be covalently attached to pharmaceutically

CC active compounds. After administration, the coupled medicament will not  
 CC diffuse from the circulating blood into the interstitium. Clearance by  
 CC liver and kidney will be kept to a minimum, ensuring a more constant  
 CC plasma level of the medicament. Suitable medicaments include those  
 CC involved in intervening blood clotting, vasodilation, function of  
 CC erythrocytes, thrombocytes and leukocytes, immune responses, and blood  
 CC levels of messenger molecules such as hormones.  
 XX  
 XX Sequence 617 AA;

Query Match 100.0%; Score 3488; DB 8; Length 617;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-207;  
 Matches 617; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPPGEPGPTGLPFPGERGSGRFPFCADGVAGPKGAGERSGPGAGKSGSPGEAGRP 60  
 Db 1 GPPGEPGPTGLPFPGERGSGRFPFCADGVAGPKGAGERSGPGAGKSGSPGEAGRP 60

QY 61 GEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPPPGARGQAGVMGFPKGA 120  
 Db 61 GEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPPPGARGQAGVMGFPKGA 120

QY 121 GEPKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 180  
 Db 121 GEPKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 180

QY 181 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 240  
 Db 181 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 240

QY 241 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 300  
 Db 241 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 300

QY 301 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 360  
 Db 301 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 360

QY 361 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 420  
 Db 361 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 420

QY 421 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 480  
 Db 421 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 480

QY 481 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 540  
 Db 481 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 540

QY 541 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 600  
 Db 541 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 600

QY 601 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 617  
 Db 601 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 617

## RESULT 2

ADQ26217  
 ID ADQ26217 standard; protein; 617 AA.  
 XX  
 XX  
 XX ADQ26217;  
 DT 23-SEP-2004 (first entry)  
 XX Human gelatine-like polypeptide Hu-3.  
 DE Human gelatine-like protein; Hu-3; microcarrier; cell culture.  
 KW Human; gelatine-like protein; Hu-3; microcarrier; cell culture.  
 XX Homo sapiens.  
 OS  
 XX

PN WO2004056976-A2.  
 XX 08-JUL-2004.  
 XX 23-DEC-2003; 2003WO-NL000922.  
 XX 23-DEC-2002; 2002EP-00080539.  
 PR (FUJF) FUJII PHOTO FILM BV.  
 XX  
 XX Bouwstra JB, Van Es AJU, Toda Y;  
 XX WPI; 2004-507711/48.  
 DR  
 XX Preparing cell culture support useful for culturing anchorage dependent  
 PT cells, involves coating microcarrier bead with gelatine or gelatine-like  
 PT protein having specific molecular weight.  
 XX  
 XX Example 1; SEQ ID NO 1; 19pp; English.  
 PS  
 XX The present sequence is the protein sequence of human recombinant gelatin  
 CC -like polypeptide Hu-3, which has a molecular weight of approximately 54  
 CC kDa and which contains multiple Gly-Xaa-Yaa triplets. In an example from  
 CC the invention, recombinant Hu-3 was immobilised on polystyrene beads  
 CC using a heterobifunctional crosslinking agent, and used in a spinner  
 CC flask cell culture. A claimed process for the preparation of a cell  
 CC culture support comprises the step of coating a microcarrier bead with  
 CC gelatine or gelatine-like protein having a molecular weight of about 40-  
 CC 200 kDa, and optionally further comprising the step of immobilising the  
 CC gelatine or gelatine-like protein on the microcarrier. In this process,  
 CC more than 75%, preferably more than 85% and more preferably more than 95%  
 CC of the gelatine or gelatine-like protein has the same molecular weight.  
 CC The gelatine or gelatine-like protein is recombinantly produced to obtain  
 CC a material of uniform molecular weight and to reduce the risk of  
 CC contamination with prions. A claimed cell support consists of microbeads  
 CC of 50-500 um size coated with a gelatine-like protein consisting of at  
 CC least 95% Gly-Xaa-Yaa triplets and containing at least 15% proline  
 CC residues and less than 5% of hydroxyproline residues, with a molecular  
 CC weight distribution showing a maximum between 40 and 200 kDa, at least  
 CC 75% of the protein molecules having a molecular weight within 2% of the  
 CC maximum. Large-scale production of expressed products can be accomplished  
 CC with gelatine-coated microcarriers.  
 XX  
 SQ Sequence 617 AA;

Query Match 100.0%; Score 3488; DB 8; Length 617;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-207;  
 Matches 617; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPPGEPGPTGLPFPGERGSGRFPFCADGVAGPKGAGERSGPGAGKSGSPGEAGRP 60  
 Db 1 GPPGEPGPTGLPFPGERGSGRFPFCADGVAGPKGAGERSGPGAGKSGSPGEAGRP 60

QY 61 GEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPPPGARGQAGVMGFPKGA 120  
 Db 61 GEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPPPGARGQAGVMGFPKGA 120

QY 121 GEPKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 180  
 Db 121 GEPKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 180

QY 181 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 240  
 Db 181 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 240

QY 241 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 300  
 Db 241 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 300

QY 301 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 360  
 Db 301 GPPGKAGRGVPGPPGAVGAPKAGDEAGAGPPGAGPAGGERGEGPAGSPFGQLPGPA 360

QY 361 GPAGERGQAGSPGPGQGLPGAGPPGAGKPGEGQVPGDLGAPGSPGAGPGPTGLP 420  
Db 361 GPAGERGQAGSPGPGQGLPGAGPPGAGKPGEGQVPGDLGAPGSPGAGPGPTGLP 420  
QY 421 GPPGERGQAGSPGPGQGLPGAGPPGAGKPGEGQVPGDLGAPGSPGAGPGPTGLP 480  
Db 421 GPPGERGQAGSPGPGQGLPGAGPPGAGKPGEGQVPGDLGAPGSPGAGPGPTGLP 480  
QY 481 GSPGSPGDKTGPAGQDGRPPGPPGAGQAGVWGFPFGKAGAGPPGKAGRGVP 540  
Db 481 GSPGSPGDKTGPAGQDGRPPGPPGAGQAGVWGFPFGKAGAGPPGKAGRGVP 540  
QY 541 GPPGAVGAGKDGAGAGQAGPPGAGPAGERGQAGSPGPGQGLPGAGPPGAGKPGEGQ 600  
Db 541 GPPGAVGAGKDGAGAGQAGPPGAGPAGERGQAGSPGPGQGLPGAGPPGAGKPGEGQ 600  
QY 601 GVPGLGAPGSPGAGG 617  
Db 601 GVPGLGAPGSPGAGG 617

RESULT 3

ADM48392  
ID ADM48392 standard; protein; 821 AA.  
XX AC  
XX ADM48392;  
XX DT 03-JUN-2004 (first entry)  
XX DE Human recombinant gelatin-like polypeptide Hu-4.  
XX KW Plasma substitute; Gelatin-like protein; plasma expander; human.  
XX OS Homo sapiens.  
XX PN EPI398324-A1.  
XX PD 17-MAR-2004.  
XX PF 11-SEP-2002; 2002EP-00078745.  
XX PR 11-SEP-2002; 2002EP-00078745.  
XX PA (FUJIF) FUJIFILM PHOTO FILM BV.  
XX PI Bouwstra JB, Toda Y;  
XX WPI; 2004-229415/22.  
XX FS Composition useful as substitute for plasma, comprises solution of saline  
PT and recombinant gelatin-like protein having colloid osmotic function.  
XX Example 1; SEQ ID NO 3; 31pp; English.

CC The present sequence is the protein sequence of recombinant gelatin-like  
CC protein Hu-3. This is a tetramer of human recombinant gelatin-like  
CC protein Hu-1. ADM48392. A claimed composition suitable as a substitute for  
CC plasma comprises a solution of saline and a protein having a colloid  
CC osmotic function. The protein is a recombinant gelatin-like protein with  
CC a molecular weight of at least 10 kDa and at most 50 kDa and an  
CC isoelectric point of less than 8. It is especially Hu-1 or Hu-deam  
CC ADM48393. The recombinant gelatin-like protein, or a dimer, trimer or  
CC tetramer of the protein, is useful as a plasma expander that has a lower  
CC clearance rate from blood circulation, provides better and predictable  
CC regulation of clearance rate and which is less susceptible to proteolytic  
CC degradation than presently used gelatin derivatives. Recombinant gelatin-  
CC like proteins that are in essence free of hydroxyproline do not give rise  
CC to an immunological reaction with blood samples containing IgE  
CC antibodies. The gelatin-like proteins can be covalently attached to  
CC pharmaceutically active compounds. After administration, the coupled  
CC medicament will not diffuse from the circulating blood into the  
CC interstitium. Clearance by liver and kidney will be kept to a minimum,  
CC ensuring a more constant plasma level of the medicament. Suitable

CC medicaments include those involved in intervening blood clotting,  
CC vasodilation, function of erythrocytes, thrombocytes and leukocytes,  
CC immune responses, and blood levels of messenger molecules such as  
CC hormones.

XX Sequence 821 AA;

Query Match 99.8%; Score 3482; DB 8; Length 821;  
Best Local Similarity 100.0%; Pred. No. 4.9e-207;  
Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPPGPPGPTGLPFPGERGSGRFPAGDAGVAGPKGAPGAGRGSPGAPGKSGPGEAGRP 60  
Db 1 GPPGPPGPTGLPFPGERGSGRFPAGDAGVAGPKGAPGAGRGSPGAPGKSGPGEAGRP 60  
QY 61 GEAGLPGAKGLTSGPSGPGDKTGPAGQDGRPPGPPGAGQAGVWGFPFGKAGAA 120  
Db 61 GEAGLPGAKGLTSGPSGPGDKTGPAGQDGRPPGPPGAGQAGVWGFPFGKAGAA 120  
QY 121 GPPGKAGRGVFPFGAGVAGPKAGQAGAGQAGPPGAGPAGRGSGQAGSPGQGLPGPA 180  
Db 121 GPPGKAGRGVFPFGAGVAGPKAGQAGAGQAGPPGAGPAGRGSGQAGSPGQGLPGPA 180  
QY 181 GPPGKAGRGVFPFGAGVAGPKAGQAGAGQAGPPGAGPAGRGSGQAGSPGQGLPGPA 240  
Db 181 GPPGKAGRGVFPFGAGVAGPKAGQAGAGQAGPPGAGPAGRGSGQAGSPGQGLPGPA 240  
QY 241 GPAGRGSPGAGPKGSPGAGRPGEAGLPGAKGLTSGPSGPGDKTGPAGQDGRP 300  
Db 241 GPAGRGSPGAGPKGSPGAGRPGEAGLPGAKGLTSGPSGPGDKTGPAGQDGRP 300  
QY 301 GPPGPPGARGQAGVWGFPFGKAGRGVFPFGAGVAGPKAGQAGAGQAGPPGPA 360  
Db 301 GPPGPPGARGQAGVWGFPFGKAGRGVFPFGAGVAGPKAGQAGAGQAGPPGPA 360  
QY 361 GPAGRGQAGSPGPGQGLPGAGPPGAGKPGEGQVPGDLGAPGSPGAGPGPTGLP 420  
Db 361 GPAGRGQAGSPGPGQGLPGAGPPGAGKPGEGQVPGDLGAPGSPGAGPGPTGLP 420  
QY 421 GPPGPPGARGQAGVAGPKAGPAGRGSPGAGPKGSPGAGRGVFPFGAGLPGAKGLT 480  
Db 421 GPPGPPGARGQAGVAGPKAGPAGRGSPGAGPKGSPGAGRGVFPFGAGLPGAKGLT 480  
QY 481 GSPGSPGDKTGPAGQDGRPPGPPGAGQAGVWGFPFGKAGAGPPGKAGRGVP 540  
Db 481 GSPGSPGDKTGPAGQDGRPPGPPGAGQAGVWGFPFGKAGAGPPGKAGRGVP 540  
QY 541 GPPGAVGAGKDGAGAGQAGPPGAGPAGERGQAGSPGPGQGLPGAGPPGAGKPGEGQ 600  
Db 541 GPPGAVGAGKDGAGAGQAGPPGAGPAGERGQAGSPGPGQGLPGAGPPGAGKPGEGQ 600  
QY 601 GVPGLGAPGSPGAG 616  
Db 601 GVPGLGAPGSPGAG 616

RESULT 4

AY84544  
ID AY84544 standard; protein; 1057 AA.

XX AC

XX AY84544;

DT 25-JUL-2000 (first entry)

XX DE A human collagen 1 (alpha1) protein helical region.

XX KW Extracellular matrix protein; self aggregation; hydroxylated proline;  
KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;  
KW collagen; fibrinogen; fibronectin; post translational hydroxylation.

XX OS Homo sapiens.

XX PF BP92586-A2.

XX PD 12-APR-2000.  
 XX PF 07-OCT-1999; 99EP-00119184.  
 XX PR 09-OCT-1998; 98US-00169768.  
 XX PA (USSU) US SURGICAL CORP.  
 XX PI Gruskin EA, Buechter DD, Zhang G, Connolly K;  
 XX WPI; 2000-259138/23.  
 DR N-PSDB; AAA12503.  
 XX  
 PT Production of extracellular matrix proteins containing 4-trans-  
 PT hydroxyproline results in native self aggregating proteins, useful on  
 PT medical implants.  
 XX  
 PS Example 10; Fig 39A-B; 260pp; English.  
 XX  
 CC The specification describes a method for producing an extracellular  
 CC matrix protein or its fragment. The extracellular matrix protein is  
 CC capable of self aggregating in a cell which does not ordinarily  
 CC hydroxylated prolines. The method comprises optimising a nucleic acid  
 CC sequence for expression in the cell by substitution of codons preferred  
 CC by that cell for naturally occurring codons not preferred by the cell;  
 CC incorporating the nucleic acid sequence into the cell; and contacting the  
 CC cell with a hypertonic growth medium containing at least one amino acid,  
 CC selected from the group consisting of trans-4-hydroxyproline and 3-  
 CC hydroxyproline to allow at least one of the amino acids to be assimilated  
 CC into the cell and incorporated into the extracellular matrix protein. The  
 CC method may be used to make host cells assimilate and incorporate trans-4-  
 CC hydroxyproline into proteins. This is especially useful in the  
 CC recombinant production of proteins such as collagen, fibrinogen and  
 CC fibronectin whose ability to self aggregate and produce functional  
 CC proteins depends on the post translational hydroxylation of proline. The  
 CC method is also useful in studying the structure and function of  
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The  
 CC present sequence represents human collagen I (alpha1) helical region,  
 CC which may be produced using the method of the invention  
 XX  
 SQ Sequence 1057 AA;

Query Match 64.1%; Score 2237.5; DB 3; Length 1057;  
 Best Local Similarity 60.2%; Pred. No. 3e-130;  
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;  
 QY 1 GPP-----GFPGTGLFPPGGERGSGRFPFGADGVAGPKPAGERGSPGPA 48  
 DB 297 GPPGAGEGKRGARGEPFTGLFPPGGERGSGRFPFGADGVAGPKPAGERGSPGPA 356  
 QY 49 GPKSGPEAGRPAGLPGAKLTGSGSGPPGDKTTPPCAGQDGRPGPPGARGQA 108  
 DB 357 GPKSGPEAGRPAGLPGAKLTGSGSGPPGDKTTPPCAGQDGRPGPPGARGQA 416  
 QY 109 GVMGFPKGAAGPPGKAGRGVPPGCAVGPAGKDGKAGAGQPPGPPGAPGARGGQGA 168  
 DB 417 GVMGFPKGAAGPPGKAGRGVPPGCAVGPAGKDGKAGAGQPPGPPGAPGARGGQGA 476  
 QY 169 GSPGFQGLPGAPPPGAGKPGGQVPGDILGAPGSGPAGE-----PGP----- 212  
 DB 477 GSPGFQGLPGAPPPGAGKPGGQVPGDILGAPGSGPAGE-----PGP----- 536  
 QY 213 -----TGLPFPGERGSGRFPFGADGVAGPKPAGERGSPGPKGSP 258  
 DB 537 GANCAPNDGAKDAGAPGAPGSGGAPLQGMFGERGAALPGKDRGDAGPKGADGSP 596  
 QY 259 GE-----AGRPGEAGLPKAK-----GLTSGSGSPGPDGKTTPPGAGQDGRP 300  
 DB 597 GKDGVRGLTGPFGPPGAPAGPDGKSGSGSPGAPGTGARGAPDGRGEPGPPGAPGAP 656  
 QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGRGKAGRGVPPGPA----- 341

DB 657 GADGQPGAKGEPGDAGAKGDAGPPGPGAGPAGPPIGNVGAPGAKGARGSPGPGATGFP 716  
 QY 342 -----VGPAGKDGEGAGQPPGAPG-----AGERGEQPGAGSPGQGLPGPAGPGEAGKP 393  
 DB 717 GAAGRVGPPGPGSGNAGPPGPPGAPGKGGKPRGETGAPRGVGGVPPGPPGAGEKGGSP 776  
 QY 394 GEQGVGDLGAPGPPGSPAGEPCTGLPGPPGERGPGSRGPPGADGVAGPKPAGERGSP 453  
 DB 777 GADGAPAGPPTGPGQIAGQGVGLPGQRCGERGFGPLPGSPGPPGKGGPSGASGERGPP 836  
 QY 454 GPAGP---KSGPGEAGRPGEAGLPGAKLTGSGSPGPDGKTTPPGAGQDGRGPPGPP 510  
 DB 837 GPMGPPGLAGPPGSGREGAFGAEGRGSDGSPGAKGDRGETGAPGPPGAPGAPGAPGV 896  
 QY 511 GARGQAGVMGPPGPKGAAGEPKKAGRGVPPGPGAVGAPGAKDGE-----AGAQ 558  
 DB 897 GPAGSGDRGETGTPAGPAGPVGPGARGPAGPQGRGDKGETGQGRGKIKGHRGFSGLQ 956  
 QY 559 GPPGP-----AGPAGERGEQPGAGSP---GFQGLPGPAGPPGAGKPGEGQVGP 603  
 DB 957 GPPGPPGSGPQSGSGAGSPAGPRGPGSAGAPGKDGGLGLPGIPGPPGRGTGDAGPV 1016  
 QY 604 GDLGAPGSPGAPG 616  
 DB 1017 GPPGPPGPPGPPG 1029  
 RESULT 5  
 ID AAY84403 standard; protein; 1058 AA.  
 XX AAY84403;  
 AC AAY84403;  
 DT 12-JUL-2000 (first entry)  
 XX Amino acid sequence of human type 1 (alpha1) collagen polypeptide.  
 DE Alpha1 collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;  
 KW collagen; mussel adhesive protein; bioadhesive.  
 XX Homo sapiens.  
 OS WO200014201-A1.  
 PN 16-MAR-2000.  
 PD 07-SEP-1999; 99WO-US020462.  
 PF 09-SEP-1998; 98US-0099652P.  
 PR (USSU) US SURGICAL CORP.  
 PA (PAOL/) PAOLELLA D N.  
 PA (GRUS/) GRUSKIN E A.  
 PA (BUEC/) BUECHTER D D.  
 XX Paolletta DN, Gruskin EA, Buechter DD;  
 XX WPI; 2000-271051/23.  
 DR N-PSDB; AA299843.  
 XX Incorporating non-natural amino acid into polypeptide, useful e.g. for  
 PT production of bioadhesives, by epoxidation or substitution of  
 PT denhydroproline residues.  
 PT Disclosure; Fig 6; 66pp; English.  
 PS The present sequence represents a human type 1 (alpha1) collagen protein.  
 XX Peptides derived from the protein were used to demonstrate incorporation  
 CC of 3,4-dehydro-L-proline into the peptide, using the method of the  
 CC invention. The specification describes a method for the incorporation of  
 CC non-natural amino acid into a polypeptide. The method comprises reacting  
 CC at least one 3,4-dehydroproline residue in the polypeptide with an  
 CC epoxidation reagent from a polypeptide containing at least one 3,4-



CC epoxypoline residue. The method is used for studying the effects of non-  
 CC natural amino acids on structure and function of polypeptides. The method  
 CC is also useful for commercial production of collagen or mussel adhesive  
 CC proteins (which are useful as bioadhesives), and for incorporating a wide  
 CC variety of groups, including therapeutic ligands and biological probes,  
 CC into polypeptides  
 XX  
 SQ Sequence 1058 AA;  
 Query Match 64.1%; Score 2237.5; DB 3; Length 1058;  
 Best Local Similarity 60.2%; Pred. No. 3e-130;  
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;  
 Qy 1 GPP-----GPGFTGLPFGPPGGRGGRGFGPGADGVAGPKGPGAGSPGPA 48  
 Db 298 GPPGAGEGKRGARGEGPTGLPFGPPGGRGGRGFGPGADGVAGPKGPGAGSPGPA 357  
 Qy 49 GPKSGPEAGRGEAGLPGAKGLTSGSPGPGDKTCTPPGAGDGRPPGPPGARGOA 108  
 Db 358 GPKSGPEAGRGEAGLPGAKGLTSGSPGPGDKTCTPPGAGDGRPPGPPGARGOA 417  
 Qy 109 GVMGFPKGAAGFPKAGRGVPGPGAVGPGAGKGEAGAGQPPGPGAGBGRGQGPA 168  
 Db 418 GVMGFPKGAAGFPKAGRGVPGPGAVGPGAGKGEAGAGQPPGPGAGBGRGQGPA 477  
 Qy 169 GSPFGQLPAGPPGAGKGEAGVPGDGLGAPGSPGAGE-----PGP----- 212  
 Db 478 GSPFGQLPAGPPGAGKGEAGVPGDGLGAPGSPGAGE-----PGP----- 537  
 Qy 213 -----TGLPFGPPGGRGGRGFGPGADGVAGPKGPGAGSPG 258  
 Db 538 GAGAPNDGAKGAGAPGAPGSGAFQLOMPGEGAAGLPGKDRGAGKAGDGP 597  
 Qy 259 GE-----AGRPGAGLPGAK-----GLTSGSPGPDGKTCTPPGAGDGRP 300  
 Db 598 GKDGVRGLTGPIGFPAGAPGDKGSGSPGAGTGAAGPDRGPPGPGAGPAGPP 657  
 Qy 301 GPPPPGARGQ-----AGVMGFPKGAAGEPKAGRGVPGPPGA----- 341  
 Db 658 GADGQPKAGPEPDAGAKGADGPPGAGPAGPPIGNVGAAPKARGAGAPPGATGFP 717  
 Qy 342 -----VSPAGKDGAGAGQPPGAGP-----AGERGEQPGSPGFGQLPFGAGPPGAGKP 393  
 Db 718 GAAGRVGPPGPGSGNAGPPGPGAGKGGKPRGCTGAGRPGVGVPPGPPGAGKGP 777  
 Qy 394 GEQGVGDLGAPGSPGAGEPGPTGLPQPPGGRGGRGFRPGADGVAGPKGPGAGRGSP 453  
 Db 778 GADGPAGAPGTPGQGIAGQGVVGLPQGRGERGFPGLPFGSPGKQPSGASGERGPP 837  
 Qy 454 GPAGP---KGSFGBAGRGEAGLPGAKGLTSGSPGPDGKTCTPPGAGDGRPPGPP 510  
 Db 838 GPMGPPGLAGPPGSGREGAFGAGSPGSDSPAKGDRGETGTGAGPFGAGPAGPVP 897  
 Qy 511 GARGQAGVMGFPKGAAGFPKAGRGVPGPGAVGPGAGKGE-----AGAQ 558  
 Db 898 GPAGKSGDRGTGTGAGPAGVPVGPAGAGPAGPQGRGDKGTGTGQGRGKGRHGFSLQ 957  
 Qy 559 GPPGP-----AGPAGERGEQGPAGSP---GFGQLPFGAGPPGAGKGEQGV 603  
 Db 958 GPPGPFSGGSGGAGPAGPPGSPAGAPKDGGLGFLPFGPPGPRGRTGDGAPV 1017  
 Qy 604 GDLGAPGSPGAG 616  
 Db 1018 GPPGPPGPPGPPG 1030  
 RESULT 6  
 ID AAY84540  
 XX AAY84540 standard; protein; 1107 AA.  
 AC AAY84540;  
 XX DT 25-JUL-2000 (first entry)

XX Amino acid sequence of a chimeric collagen 1 (alpha1)/decorin protein.  
 DE Extracellular matrix protein; self aggregation; hydroxylated proline;  
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;  
 KW collagen; fibrinogen; fibronectin; post translational hydroxylation;  
 KW decorin; chimera.  
 XX Homo sapiens.  
 OS Unidentified.  
 OS Chimeric.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 858 /note= "Gly encoded by GCT"  
 FT  
 PN EP92586-A2.  
 XX  
 PD 12-APR-2000.  
 XX  
 PF 07-OCT-1999; 99EP-00119184.  
 XX  
 PR 09-OCT-1998; 98US-00169768.  
 XX  
 PA (USSU ) US SURGICAL CORP.  
 XX  
 XX Gruskin EA, Buechter DD, Zhang G, Connolly K;  
 PI WPI; 2000-259138/23.  
 DR N-PSDB; AAA12500.  
 XX  
 PT Production of extracellular matrix proteins containing 4-trans-  
 PT hydroxyproline results in native self aggregating proteins, useful on  
 PT medical implants.  
 XX  
 XX Claim 24; Fig 18; 260pp; English.  
 XX  
 CC The specification describes a method for producing an extracellular  
 CC matrix protein or its fragment. The extracellular matrix protein is  
 CC capable of self aggregating in a cell which does not ordinarily  
 CC hydroxylated prolines. The method comprises optimising a nucleic acid  
 CC sequence for expression in the cell by substitution of codons preferred  
 CC by that cell for naturally occurring codons not preferred by the cell;  
 CC incorporating the nucleic acid sequence into the cell; and contacting the  
 CC cell with a hypertonic growth medium containing at least one amino acid,  
 CC selected from the group consisting of trans-4-hydroxyproline and 3-  
 CC hydroxyproline to allow at least one of the amino acids to be assimilated  
 CC into the cell and incorporated into the extracellular matrix protein. The  
 CC method may be used to make host cells assimilate and incorporate trans-4-  
 CC hydroxyproline into proteins. This is especially useful in the  
 CC recombinant production of proteins such as collagen, fibrinogen and  
 CC fibronectin whose ability to self aggregate and produce functional  
 CC proteins depends on the post translational hydroxylation of proline. The  
 CC method is also useful in studying the structure and function of  
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The  
 CC present sequence represents a chimeric collagen 1 (alpha1)/decorin  
 CC protein, which may be produced using the method of the invention  
 XX  
 SQ Sequence 1107 AA;  
 Query Match 64.1%; Score 2237.5; DB 3; Length 1107;  
 Best Local Similarity 60.2%; Pred. No. 3.1e-130;  
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;  
 Qy 1 GPP-----GPGFTGLPFGPPGGRGGRGFGPGADGVAGPKGPGAGSPGPA 48  
 Db 297 GPPGAGEGKRGARGEGPTGLPFGPPGGRGGRGFGPGADGVAGPKGPGAGSPGPA 356  
 Qy 49 GPKSGPEAGRGEAGLPGAKGLTSGSPGPGDKTCTPPGAGDGRPPGPPGARGOA 108  
 Db 357 GPKSGPEAGRGEAGLPGAKGLTSGSPGPGDKTCTPPGAGDGRPPGPPGARGOA 416  
 Qy 109 GVMGFPKGAAGFPKAGRGVPGPGAVGPGAGKGEAGAGQPPGPGAGBGRGQGPA 168



717	Db	GAAGRIVGPPGSGNAGFPQPPGGPAGKGGKGRGETGPAGRPGVEVPPGPPGPAGEKGP	776
394	QY	GEQVFPDGLGAPGSPGAPGPTGLPGPFGERGGPSRGFPAGDVAGPKGPAGERGSP	453
777	Db	GADGPAGAPCTPGPQGIAGQGVVGLPGQRGERGFPCLPGPSCGPKQKQSGASGERGPP	836
454	QY	GRAGP---KGSPEAGRPGEAGLPGAAGLGTGSPGCPDGKTCPPGPAGQDGRGPPGPP	510
837	Db	GPNGPPLGAGPPGESGREGAPGAEGRDQSGCAKGRGETGPAGPAGPAGPAGPVP	896
511	QY	GARGQAGVMGFPFGKAAGEFGKAGRGVPGPPCAVGPAGKQGE-----AGAQ	558
897	Db	GPAGKSGDRGETGTPAGPAGVGPAGARGPAGPQGPGRGDKGETGEQSDRGIKHGRGSGLG	956
559	QY	GPFGP-----AGPAGRGEGCPAGSP---GFQGLPGPAGPPGAGKPGGEOGVP	603
957	Db	GPFGPPGSGPGEQFQSGNAGPAGPRGPPGSGAGACKQGLNGLPGFIPGPPRGRTGDAGVP	1016
604	QY	GDLAGPSPGSPAG	616
1017	Db	GPFGPPGPPGPPG	1029

RESULT 8  
AAW68485  
ID AAW68485 standard; protein: 1464 AA.

AC  
XX  
XX  
DT  
XX  
XX  
DE  
AAW69485;  
08-DEC-1998 (first entry)  
Human recombinant collagen protein.

XX Primer; PCR; amplification; human; collagen; mammal; plant; prosthesis;  
KW cardiac valve; ligament; tendon; skin; gingival implant; perfumes;  
KW nerve regeneration; antibiotic; growth factor; cancer; inflammatory;  
KW gelatin; glue; food

xx	OS	Synthetic.
xx	OS	Homo sapiens.

XX	Key	Location/Qualifiers
FH	Peptide	1..22
FT		/note= "signal peptide"
FT	Protein	23..999
FT		/note= "mature protein"
FT	Cleavage-site	161
FT		/note= "cleavage site for aminopeptidase"
FT	Cleavage-site	1218
FT		/note= "cleavage site for carboxypeptidase"

DN	WO9827202-A1.	
XX		
XX	25-JUN-1998.	
XX		
XX	17-DEC-1997;	97WO-FR002331.
XX		
XX	17-DEC-1996;	96FR-00016224.
XX		
XX	(BIOC-) BIOCEM SA.	
PA		

XX Gruber V, Exposito J, Ruggiero F, Conte J, Garrone R, Merot B;  
PI Bournat P;  
XX WPI: 1998-362771/31.  
XX N-PSDR: AAV60814.  
DE

xx New recombinant nucleic acid for expressing collagen or derivatives in plants - useful as, e.g. bio-materials and in therapeutic, cosmetic and odontological compositions.

XX PS Disclosure; Fig 7; 138pp; French.

The invention relates to the production of mammalian collagen in plants. 2 clones: alpha3 and alpha2, spanning the human collagen type I gene were isolated from a MG-63 osteosarcoma library. Clone alpha3 contained 83 bp of the 5' untranslated region and the first 1920 bp of coding sequence, whereas clone alpha22 contained sequence encoding amino acids 171-1454 of the protein and around 500 bp of the 3' untranslated region. The 2 clones were used to generate a number of fragments which were used to construct composite sequences encoding variant collagen molecules. The fragments are: (A) containing nucleotides (nt) -4 to 479; (B) containing TAA upstream of the sequence encoding the PRS (pathogenesis-related protein S) signal peptide and bases 66-77 from the sequence encoding the N-terminus of the pro-collagen amino propeptide domain; (C) the whole of the amino propeptide domain (nt 72-479); (D) all of the amino-telopeptide domain (nt 474-534) and the N-terminus of the helicaloid region (nt 535-1920); (E) the DraIII-BamHI fragment (1709-2808) of alpha22, encoding aa 567-936 of the central helicaloid domain; (F) the BamHI-EcoRI (2803-4362) region of alpha22, encoding aa 936-1192 in the central helicaloid domain and aa 1193-1454 in the C-propeptide domain; (G) the C-terminus of the C-propeptide domain (aa 1346-1464) plus stop codons, and (H) as G but encoding aa 1343-1401 and also including the KDEU motif for retention in the ER. This sequence represents a recombinant human collagen. The encoding gene was constructed from fragments (A), (D), (E), (F) and (G). The recombinant gene is used for expression of mammalian collagen in plant cells. The transformed plants, their extracts and parts are useful as biomaterials (haemostatic compresses, sponges or bandages) and in pharmaceutical, medical, odontological, cosmetic and biotechnological compositions (e.g. as prostheses for cardiac valves, ligaments or tendons; skin substitutes; gingival implants; microcapsules for perfumes; guide tubes for nerve regeneration; slow release products for antibiotics, growth factors, anticancer agents or anti-inflammatory; surgical thread and components of cements). They are suitable for treating any disorder related to collagen dysfunction and gelatin, produced from collagen, is used to produce glues, surgical prostheses and foods

Sequence 1464 AA:

Query Match	64.1%	Score 2237.5;	DB 2;	Length 1454;
Best Local Similarity	60.2%	Pred. No. 3.9e-130;		
Matches 441:	Conservative	23;	Mismatches 152;	Indels 117;
	Gaps			12;

Qy	1	GPP-----GEPQPTGLPGPPGRRGGPSGRGFGADGVAGPKGPAGRRGSGPPA	48
Db	458	GPFGPAGEKRGARCEPPTGLPGPPGRRGGPSGRGFGADGVAGPKGPAGRRGSGPPA	517
Qy	49	GPKSGPCEAGRPEAGLPGAAGLGTGSPGSGPPDGKTGPPGPAQODGRPGPPGPPARGQA	108
Db	518	GPKSGPCEAGRPEAGLPGAAGLGTGSPGSGPPDGKTGPPGPAQODGRPGPPGPPARGQA	577
Qy	109	GVNMGFPQKGAAGEPGKAGRGVPPGPAVGPAGKDGEAGAAQOQPPGPAGPAGRGEGOGPA	168
Db	578	GVNMGFPQKGAAGEPGKAGRGVPPGPAVGPAGKDGEAGAAQOQPPGPAGPAGRGEGOGPA	637
Qy	169	GSPGFQGLPGPAPPGPEAGKPGGQGVPGDLGAGPSPGAGE-----PGP----	212
Db	638	GSPGFQGLPGPAPPGPEAGKPGGQGVPGDLGAGPSPGAGE-----PGP----	697
Qy	213	-----TGLPGPPGRRGGFGSGRFGFGADGVAGPKGPAGRRGSGPPAGPKGSP	258
Db	698	GANGAPNDGAKGDAGAPGAPGSGQAGPLQGMTPGERGAALPGPKGDRGAGPKGADGSP	757
Qy	259	GE-----AGRPGEAGLPGAK-----GLTSGPSGPGDKTGPAGAGQDGRP	300
Db	758	GKDQVRLGTGPIGPPGPAGAPGDKSGSGPSGPGAGPTGARGAPGDRGPFPPGPAAGFAGPP	817
Qy	301	GPFGPPGARQQ-----AGVMGFTPGKGAAGEPGKAGRGVPPGPGA-----	341
Db	818	GADQPCAKGCEPDAGAKGDAGPPGPAGPAGPPGPIGNVGAPKAGARGAGPPGATGFP	877
Qy	342	-----VQPAKDEAGAGQPPGAPG---AGERGEQGPAGSPGQGLPGPAGPPGPEAGKP	393
Db	878	GAAGRVPFPSPGNAGPFGPPGPAKGGKGPGETGAPRGPEVBVPPPPGPPAGRGKSP	937
Qy	394	GEQVPGDLAGPSPGAPGPGTGTGLPGPGEERGSGRFPFGADGVAGPKGPAGRRGSP	453

Db 938 GADGAPAGTPTGQIAGQWGLPGQGERGPGGLPGSPGKQKQPSGASGERGPP 997  
 QY 454 GPAGP---KGSPEAGRPEAGLPCAKGLTSGSPGPDGKTGPPGAGDGRPGPP 510  
 Db 998 GPMGPPGLAGPPEESGREGAPGAEPSGDRGSPGAKGDRGTGTGAGPPGAPGAPGPV 1057  
 QY 511 GARGQAGVMGPPGPKAAGEPCKAGRGVPGPPGAVGPAKDGK-----AGAQ 558  
 Db 1058 GPAGKSGDRGTGTGAPGAPGVPAGARGPAGPQGRGKGTGQDRIKGRHGSGLQ 1117  
 QY 559 GPPGP-----AGPAGERGQCPAGSP---GQGLPGAPGPPGKRGEGQV 603  
 Db 1118 GPPGPPGSGGQPSGASGAPRPGPSAGAPGKQGLNGLPGIPGPRGRGTGDGPV 1177  
 QY 604 GDLGAPGSPGAP 616  
 Db 1178 GPPGPPGPPGPPG 1190

## RESULT 9

ADD45059  
 ID ADD45059 standard; protein; 1464 AA.

AC ADD45059;

DT 29-JAN-2004 (first entry)

DE Human Protein P02452, SEQ ID NO 10491.

KW Human; pain; neuronal tissue; gene therapy;

KW spinal segmental nerve injury; chronic constriction injury; CCI;

KW spared nerve injury; SNI; Chung.

OS Homo sapiens.

PN WO2003016475-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-US025765.

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

PA (GEHO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

XX GENBANK; P02452.

XX New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a

CC method for identifying a compound or small molecule that regulates the CC activity in an animal of one or more of the polypeptides given in the CC specification, a method for identifying a compound useful in treating CC pain and a pharmaceutical composition comprising the one or more CC polypeptides or their antibodies. The polynucleotide or the compound that CC modulates its activity is useful for preparing a medicament for treating CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene CC therapy). The sequence presented is a human protein (shown in Table 2 of CC the specification) which is differentially expressed during pain. Note: CC the sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic form directly from WIPO at CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1464 AA;

Query Match 64.1%; Score 2237.5; DB 7; Length 1464;  
 Best Local Similarity 60.2%; Pred. No. 3.9e-130;  
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPQPPGERGPGSRGPPGADGVAGPKPAGERGSPGA 48  
 Db 458 GPPGAGEGKRGARGEPGPTGLPQPPGERGPGSRGPPGADGVAGPKPAGERGSPGA 517  
 QY 49 GPKGSPGABRPEAGLPCAKGLTSGSPGPDGKTGPPGAGDGRPGPPGARGQA 108  
 Db 518 GPKGSPGABRPEAGLPCAKGLTSGSPGPDGKTGPPGAGDGRPGPPGARGQA 577  
 QY 109 GVMGPPGPKAAGEPCKAGRGVPGPPGAVGPAKDGAGAGQPPGAPAGERGEOGPA 168  
 Db 578 GVMGPPGPKAAGEPCKAGRGVPGPPGAVGPAKDGAGAGQPPGAPAGERGEOGPA 637  
 QY 169 GSPGQGLPGAPGPEAGKTPGQVCDLGAPOFSGPAGE-----PGP--- 212  
 Db 638 GSPGQGLPGAPGPEAGKTPGQVCDLGAPOFSGPAGE-----PGP--- 697  
 QY 213 -----TGLPSPGEGSGSGSRGPPGADGVAGPKPAGERGSPGAPGKSP 258  
 Db 698 GANGAPGNDGAKGDAGAPGSGQAPGLQMPGERGAAGLPGPKGDRGDAGPGADGSP 757  
 QY 259 GE-----AGRPGEAGLPGAK-----GLTSPGSPGPDGKTGPPGAGDGRP 300  
 Db 758 GKDGVRLTGLTGPPEGPAGAPGDKGESGSPGAPGTGARGAPGDRGPPGPPGAPGAP 817  
 QY 301 GPPGPPGARGO-----AGVMGPPGPKAAGEPCKAGRGVPGPPGARGQA 341  
 Db 818 GADQPGAKGEGFDAGAKGDAGPFPAGPAGPFPPIGNVGAAPGAKGARGSAGPFGATGFP 877  
 QY 342 -----VGPAGKDGAGAGAQPPGAPG---AGERGQPPGSPGQGLPGAPGPPGAGK 393  
 Db 878 GAAGRVGPPGSGNAGPPGPPGAPGKGGKGRGTGTGAPRGVEVPPGPPGAPGKSP 937  
 QY 394 GEQVPGDLGAPGSPGAPGEPGPTGLPQPPGERGPGSRGPPGADGVAGPKPAGERGSP 453  
 Db 938 GADGAPAGPTGPGQIAGRGVVLPGQRGGERGPPGLPGSPGPPGKQSGSGASGERGPP 997  
 QY 454 GPAGP---KGSPPGABRPEAGLPCAKGLTSGSPGPDGKTGPPGAGDGRPGPPGPP 510  
 Db 998 GPMGPPGLAGPPEESGREGAPGAEPSGDRGSPGAKGDRGTGTGAGPPGAPGAPGPV 1057  
 QY 511 GARGQAGVMGPPGPKAAGEPCKAGRGVPGPPGAVGPAKDGK-----AGAQ 558  
 Db 1058 GPAGKSGDRGTGTGAPGAPGVPAGARGPAGPQGRGKGTGQDRIKGRHGSGLQ 1117  
 QY 559 GPPGP-----AGPAGERGEOGAPGSP---CFQGLPGAPGPPGKRGEGQV 603  
 Db 1118 GPPGPPGSGGQPSGASGAPRPGPSAGAPGKQGLNGLPGIPGPRGRGTGDGPV 1177  
 QY 604 GDLGAPGSPGAP 616  
 Db 1178 GPPGPPGPPGPPG 1190





PT vaccine, and for treating autoimmune disorders, infections and cancer.  
PS Example 3; Fig 8; 168pp; English.  
XX  
CC The present sequence is porcine alpha(I) collagen. The present invention  
CC relates to recombinant synthesis of collagens and gelatins derived from  
CC animals. Collagen is useful in medical, pharmaceutical, food and cosmetic  
CC industries. Collagen is an important component of arterial sealants, bone  
CC grafts, drug delivery system, dermal implants, haemostats, and  
CC incontinence implants, and for treating autoimmune disorders such as  
CC rheumatoid arthritis. Collagen is useful in food products such as sausage  
CC casings, and in cosmetics or facial and skin products such as  
CC moisturisers. Recombinant gelatin is useful in vaccine formulations for  
CC treating viral infections, autoimmune diseases and cancer. Gelatin is  
CC useful in the manufacture or as a component of various pharmaceutical and  
CC medical devices and products, in food and beverage industries, in hair  
CC care and skin care products, as a glue or adhesive in various  
CC manufacturing processes, as a light-sensitive coating in various  
CC electronic devices, as photoresist base in photolithographic processes,  
CC in printing and photographic applications, in laboratory application, and  
CC as a component in various gels used for biochemical and electrophoretic  
CC analysis, including enzymographic gels  
XX  
SQ Sequence 1449 AA;

Query Match 64.0%; Score 2234; DB 4; Length 1449;  
Best Local Similarity 62.0%; Pred. No. 6.4e-130;  
Matches 438; Conservative 21; Mismatches 154; Indels 94; Gaps 13;  
QY 1 GPPGPTGTPGPPGGRGGSGFPCADGAVGPKPAGERSGPPGAGKSGPGEAGRP 60  
Db 472 GARGEPPAGLPFGPPGGRGGSGFPCADGAVGPKPAGERSGPPGAGKSGPGEAGRP 531  
QY 61 GEAGLPKAGLTGSPGPPGDKTGPAGQDGRPPGPPGARGQAGVGMGPPGKGA 120  
Db 532 GEAGLPKAGLTGSPGPPGDKTGPAGQDGRPPGPPGARGQAGVGMGPPGKGA 591  
QY 121 GEPKAGRGVPPGAVGPKAGKGEAGACQPPGPPGAGERSGPPGAGKSGPGEAGRP 180  
Db 592 GEPKAGRGVPPGAVGPKAGKGEAGACQPPGPPGAGERSGPPGAGKSGPGEAGRP 651  
QY 181 GPPGAGKRGVPPGDLGAPGSPGAGE-----PGP----- 212  
Db 652 GPPGAGKRGVPPGDLGAPGSPGAGE-----PGP----- 711  
QY 213 --TGLPDPGPPGGRGGSGFPCADGAVGPKPAGERSGPPGAGKSGPGE-----A 261  
Db 712 GDAGAPAGSQGAPGLQGMPPGGRGAAGLPKPKDGRGDPAGPKADGAPGKGLTGP 771  
QY 262 GRPGEAGLPKAGLTGSPGPPGDKTGPAGQDGRPPGPPGARGQAGVGMGPPGK 317  
Db 772 GPPGAGKRGVPPGDLGAPGSPGAGE-----PGP----- 831  
QY 318 --PQKGAAGEPKKAGRGVPPGPA-----VGPAGKGEAGACQPPGPPGAGE 365  
Db 832 TGPPGPIGSGVAPGPKAGRGAGFPATGFPAGAGRVGPPGPPGAGKSGPGEAGRP 891  
QY 366 ---RGEQPPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 422  
Db 892 KGRPGETPAGPGEAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 951  
QY 423 PGEKGGPSRGPAGDGVAGPKPAGERSGPPGAGKSGPGEAGLPKAGLTGS 482  
Db 952 RGERGPPGLPSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1008  
QY 483 P---GSPGPD---GKTGPPGAGQDGRPPGPPGPPGPPGPPGPPGPPGPPGPPG 536  
Db 1009 PGRDCAFPKDRGESGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1068  
QY 537 RGVPGPAGVPAKQGE-----AGAGQPPPP-----AGAPGERGE 572  
Db 1069 RPAAGPQGRDKGETGEQDGRGTHKGRGSLGQPPGPPGPPGPPGPPGPPGPPG 1128

QY 573 QGPAGSP---GFGGLPGPAGPPGAGKPGCEQGVPGDLGAPGSPGAG 616  
Db 1129 PGAGAGPGKGLNGLPGPIGPPGRTGCDAGVPGPDPGPPGPPG 1175  
RESULT 13  
ADE87050  
ID ADE87050 standard; protein; 1161 AA.  
XX ADE87050;  
AC ADE87050;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
XX Human pancreatic cell protein sequence SeqID510.  
XX  
XX neoplastic pancreatic cell; pancreatic cancer; pancreatic cancer;  
XX cancer death; cytostatic; vaccine; Gene therapy;  
XX non-cancerous pancreas disease; human.  
XX Homo sapiens.  
XX  
PN WO2003060145-A2.  
XX  
XX 24-JUL-2003.  
XX  
XX 19-DEC-2002; 2002WO-US040655.  
XX  
XX 21-DEC-2001; 2001US-0342768P.  
XX  
XX (DIAD-) DIADEXUS INC.  
XX  
XX Sun Y, Liu C;  
XX  
XX WPI; 2003-587286/55.  
XX  
XX N-PSDB; AD387387.  
XX  
XX New pancreatic specific nucleic acid molecule or protein for diagnosing,  
XX staging, imaging, monitoring, preventing or treating pancreatic cancer or  
XX non-cancerous disease states of the pancreas.  
XX  
XX Claim 12; SEQ ID NO 510; 635pp; English.  
XX  
XX This invention relates to novel nucleic acids and proteins present in  
XX normal and neoplastic pancreatic cells. Pancreatic cancer is a common  
XX cause of cancer death worldwide, therefore accurate methods of diagnosis  
XX and treatment are required. Compounds which modulate the proteins of the  
XX invention may have cytostatic activity and the protein and DNA sequences  
XX of the invention may be useful for the development of a vaccine or in  
XX gene therapy. The composition and methods are useful in diagnosing,  
XX staging, imaging, monitoring, preventing or treating pancreatic cancer  
XX and non-cancerous disease states of the pancreas. The present sequence is  
XX that of a human pancreatic protein of the invention.  
XX  
SQ Sequence 1161 AA;  
Query Match 64.0%; Score 2233.5; DB 7; Length 1161;  
Best Local Similarity 60.0%; Pred. No. 5.7e-130;  
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;  
QY 1 GPP-----GEPGPTGLPQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 48  
Db 155 GPPGAGBEGKRGAGEPPTGLPQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 214  
QY 49 GPKGSPGAGRGGEAGLPGAKLTGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 108  
Db 215 GPKGSPGAGRGGEAGLPGAKLTGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 274  
QY 109 GYWGPPGPKGAAGEPKKAGRGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 168  
Db 275 GYWGPPGPKGAAGEPKKAGRGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 334  
QY 169 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 212

Db 335 GSPGQGLPGPAGPPGAGKPEQGVFDLGPAGSPGARGGPPGGERGVQGGPPGAPGR 394  
 QY 213 -----TGLPGPGERGGSGRFFGADGVAGPKGPAGRGSPGAGKGPSP 258  
 Db 395 GAGAPGNDGAKGADAGAFGAFSGQAFGLQMPGREGAAGLPKPKGDRGDAGPKGADGSP 454  
 QY 259 GE-----AGRPGEAGLPGAK-----CLTSPGSPGPDGKTGPPGAGQDGRP 300  
 Db 455 GKDVGRGLTGIPGPPGAGAPGDKESGSPGAPGTGARGAPCDRGEPGPPGAGPAP 514  
 QY 301 GPPGPPGARGO-----AGVMGFPKGAAGEPKKAGRGVGPVGPAA---- 341  
 Db 515 GADGQKAGKGPBGDAGAKGADAGPPGAPGAPGPPGFIQNVGAPKAGARGAGPPGATGFP 574  
 QY 342 ----VGPAGKDGAGAGPPGAPG---AGERGEQGPAGSPGQGLPGPAGPPGAGKP 393  
 Db 575 GAARVGPSPGSGNAGPPGPPGAPGKGGKGRGTGAPRGPEVGPVGPAGGKGPSP 634  
 QY 394 GEQGVGDGLGAPSPGAPGPGTGLPFPGERGGPSRFFPGADGVAGPKGPAGRGSP 453  
 Db 635 GADGAPAGPPTPGQIAGQGVVGLPQGRGERGFPGLPGSGEPKQKQSPGASGERGPP 694  
 QY 454 GPAGP---KSPGACRPGAGLPGAKLTGSPGSPGPDGKTGPPGAGQDGRPPGPP 510  
 Db 695 GPMGPPGLAGPPGSGREGAPGAEPSGRDGPAGKGDRTGAPGPPGAPGAPGPPV 754  
 QY 511 GARGQAGVMGPPGKGAAGPPGKAGRGVFPVGVGAPGAPGKDG-----AGAQ 558  
 Db 755 GPAGKSGDRGTGAPGAPGVVGVGAPGAPGQPCDKGTGEGQDRIKGRHSGSLQ 814  
 QY 559 GPPGP-----AGPAGERGEQGPAGSP---GFQGLPGPAGPPGAGKYPGQGV 603  
 Db 815 GPPGPPGSPGQSPGSGAGPAPRGPPGSGAPGKDGKGLNGLPGIPGPPRGRGTGAGPV 874  
 QY 604 GDLGAPGSPGAPG 616  
 Db 875 GPPGPPGPPGPPG 887

## RESULT 14

ABG93947  
 ID ABG93947 standard; protein; 1461 AA.

AC ABG93947;

DT 26-NOV-2002 (first entry)

DE Human polypeptide orthologous to DACC-11.

KW Human; deer; rat; mouse; DACC; deer antler cartilage cell;  
 KW cell stimulation; cell inhibition; cell growth; cell division;  
 KW mesenchymal cell; chondrocyte; chondrogenesis; osteogenesis; growth;  
 KW repair; regeneration; restoration; extracellular matrix;  
 KW cartilaginous matrix; cartilage; disc; connective tissue; agonist;  
 KW antagonist; gene therapy.

OS Homo sapiens.

FN WO200264625-A1.

FD 22-AUG-2002.

PF 15-FEB-2002; 2002WO-AU000163.

XX 15-FEB-2001; 2001AU-00003116.

XX (ADFP-) ADP PHARM PTY LTD.

PA (UNSY ) UNIV SYDNEY.

XX Roubin R, Ghosh P;

XX WPI; 2002-643456/69.

XX

PT Stimulating or inhibiting cell growth and/or division, useful for  
 PT stimulating chondrogenesis, cartilage, disc or connective tissue growth,  
 PT repair, and/or regeneration, comprises administering deer antler  
 PT cartilage gene.

XX Claim 13; Page 154-160; 214pp; English.

XX The invention discloses a method for stimulating or inhibiting cell  
 CC growth and/or division which comprises contacting or inserting into an  
 CC animal cell a polypeptide comprising one of the deer antler cartilage  
 CC cell (DACC) clones disclosed. More particularly, the method relates to  
 CC these polypeptides stimulating mesenchymal cell growth and/or division  
 CC and to transfecting these cells and chondrocytes with vectors carrying  
 CC the genes of these polypeptides capable of stimulating chondrogenesis,  
 CC osteogenesis, growth, repair, regeneration and/or restoration of the  
 CC extracellular matrix. The chondrocytes selectively express genes required  
 CC to form a cartilaginous matrix. The DACC polypeptides and polynucleotides  
 CC are useful for identifying an agent that modulates the activity of the  
 CC polypeptide, for stimulating mesenchymal cell growth and/or division by  
 CC exposing animal mesenchymal cells to conditioned media or its active  
 CC fraction, obtained from deer antler cartilage cells, for inhibiting cell  
 CC growth and/or division by inserting into an animal cell, a compound which  
 CC inhibits the translation of the polynucleotide encoding the DACC. The  
 CC method and the polypeptides are useful for stimulating mesenchymal cell  
 CC growth and/or division or for stimulating chondrogenesis, cartilage, disc  
 CC or connective tissue growth, repair, regeneration and/or restoration in  
 CC an animal. The polynucleotides, polypeptides, agonists and antagonists  
 CC may be used in treatment modalities, specifically in gene therapy. The  
 CC polypeptides can be used as bait proteins in two- or three-hybrid assay  
 CC to identify other proteins, which bind to or interact with the  
 CC polypeptide and are involved in modulating cell growth and/or division.  
 CC The sequences presented in ABG93923-ABG93948 are the proteins encoded by  
 CC the DACC cDNA clones

XX Sequence 1461 AA;

QY Query Match 64.0%; Score 2233.5; DB 5; Length 1461;  
 Best Local Similarity 60.0%; Pred. No. 6.9e-130;  
 Matches 440; Conservative 23; Mismatches 133; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLFPPGERGGPSRFFPGADGVAGPKGPAGRGSPGPA 48  
 Db 455 GPPGAPGEGKRGARGEPGPTGLFPPGERGGPSRFFPGADGVAGPKGPAGRGSPGPA 514  
 QY 49 GPKGSPGAGRPGEAGLPGAKLTGSPGSPGPDGKTGPPGAGODGRPPGPPGARGQA 108  
 Db 515 GPKGSPGAGRPGEAGLPGAKLTGSPGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 574  
 QY 109 GVMGPPGKGAAGEPGKAGRGVFPVGVGAPGAKDGEAGAQGPPGAPGAGERGEGCPA 168  
 Db 575 GVMGPPGKGAAGEPGKAGRGVFPVGVGAPGAKDGEAGAQGPPGAPGAGERGEGCPA 634  
 QY 169 GSPGQGLPGPAGPPGKAGKGEQGVQDGLGAPGSPGAPG-----RGP----- 212  
 Db 635 GSPGQGLPGPAGPPGKAGKGEQGVQDGLGAPGSPGARGRGVFPGPPGAPGR 694  
 QY 213 -----TGLPFPGERGGPSRFFPGADGVAGPKGPAGRGSPGAPGKGPSP 258  
 Db 695 GAGAPGNDGAKGADAGAFGAFSGQAFGLQMPGREGAAGLPKPKGDRGDAGPKGADGSP 754  
 QY 259 GE-----AGRPGEAGLPGAK-----CLTSPGSPGPDGKTGPPGAGQDGRP 300  
 Db 755 GKDVGRGLTGIPGPPGAGAPGDKESGSPGAPGTGARGAPCDRGEPGPPGAGPAP 814  
 QY 301 GPPGPPGARGO-----AGVMGFPKGAAGEPKKAGRGVGPVGPAA---- 341  
 Db 815 GADGQKAGKGPBGDAGAKGADAGPPGAPGAPGPPGFIQNVGAPKAGARGAGPPGATGFP 874  
 QY 342 ----VGPAGKDGAGAGPPGAPG---AGERGEQGPAGSPGQGLPGPAGPPGAGKP 393  
 Db 875 GAARVGPSPGSGNAGPPGPPGAPGKGGKGRGTGAPRGPEVGPVGPAGGKGPSP 934  
 QY 394 GEQGVGDGLGAPSPGAPGPGTGLPFPGERGGPSRFFPGADGVAGPKGPAGRGSP 453





Tue Oct 19 09:27:57 2004

us-10-658-989a-2.rag

Page 14

Job time : 104.398 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:31:20 ; Search time 24.4569 Seconds  
(without alignments)  
1673.075 Million cell updates/sec

Title: US-10-658-989A-2

Perfect score: 3488  
Sequence: 1 GPPGEGPTCLPFGPGERG.....GEGVFDLGAQPGSPAGG 617

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A COMB.pap.\*  
2: /cgn2\_6/ptodata/1/iaa/5B COMB.pap.\*  
3: /cgn2\_6/ptodata/1/iaa/6A COMB.pap.\*  
4: /cgn2\_6/ptodata/1/iaa/6B COMB.pap.\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS COMB.pap.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2237.5	64.1	1464	4	US-09-331-347C-21
2	2228.5	63.9	1341	3	US-08-963-825-18
3	2228.5	63.9	1341	3	US-09-500-811-18
4	2228.5	63.9	1341	3	US-09-570-573-18
5	2228.5	63.9	1341	3	US-09-548-608-18
6	2224.5	63.8	1461	4	US-09-585-887-9
7	2224.5	63.8	1461	4	US-09-289-578-9
8	2208.5	63.3	1057	3	US-08-931-820-1
9	2191	62.8	822	3	US-09-219-849-49
10	1955	56.0	1017	4	US-08-468-996-10
11	1943	55.7	1060	3	US-08-931-820-3
12	1943	55.7	1418	3	US-08-963-825-20
13	1943	55.7	1418	3	US-09-010-999-1
14	1943	55.7	1418	3	US-09-500-811-20
15	1943	55.7	1418	3	US-09-570-573-20
16	1943	55.7	1418	3	US-09-548-608-20
17	1936	55.5	1442	2	US-08-316-650-12
18	1936	55.5	1442	5	PCT-US95-02251-12
19	1882.5	54.0	595	3	US-09-219-849-48
20	1882.5	54.0	595	3	US-09-219-849-50
21	1882.5	54.0	1064	1	US-08-642-255-62
22	1853	53.1	1057	3	US-08-931-820-4
23	1837	52.7	720	3	US-09-219-849-4
24	1837	52.7	777	1	US-08-642-255-53
25	1833.5	52.6	1078	3	US-08-963-825-21
26	1833.5	52.6	1078	3	US-09-500-811-21
27	1833.5	52.6	1078	3	US-09-570-573-21

28 1833.5 52.6 1078 3 US-09-548-608-21 Sequence 21, Appli  
29 1764 50.6 1024 3 US-08-931-820-2 Sequence 2, Appli  
30 1764 50.6 1366 3 US-08-963-825-19 Sequence 19, Appli  
31 1764 50.6 1366 3 US-09-500-811-19 Sequence 19, Appli  
32 1764 50.6 1366 3 US-09-570-573-19 Sequence 19, Appli  
33 1764 50.6 1366 3 US-09-548-608-19 Sequence 19, Appli  
34 1764 50.6 1366 3 US-09-585-887-10 Sequence 10, Appli  
35 1762 50.5 1366 4 US-09-289-578-10 Sequence 10, Appli  
36 1699.5 48.7 1065 1 US-08-642-255-72 Sequence 72, Appli  
37 1674 48.0 960 3 US-09-219-849-5 Sequence 5, Appli  
38 1645 47.2 492 4 US-08-468-996-12 Sequence 12, Appli  
39 1644 47.1 1806 4 US-09-919-497-56 Sequence 56, Appli  
40 1619.5 46.4 829 1 US-08-642-255-132 Sequence 132, App  
41 1619.5 46.4 829 1 US-08-397-633A-53 Sequence 53, Appli  
42 1619.5 46.4 837 1 US-08-175-155-68 Sequence 68, Appli  
43 1619.5 46.4 837 1 US-08-477-509B-103 Sequence 103, App  
44 1619.5 46.4 837 1 US-08-642-255-101 Sequence 101, App  
45 1619.5 46.4 837 2 US-08-707-237A-75 Sequence 75, Appli

ALIGNMENTS

RESULT 1

US-09-331-347C-21  
; Sequence 21, Application US/09331347C  
; Patent No. 6617431  
; GENERAL INFORMATION:  
; APPLICANT: Meristem Therapeutics, S.A.  
; APPLICANT: Meristem Therapeutics, S.A.  
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, and  
; FILE REFERENCE: 1149-3  
; CURRENT APPLICATION NUMBER: US/09/331,347C  
; CURRENT FILING DATE: 1999-08-17  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 21  
; LENGTH: 1464  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-331-347C-21

Query Match 64.1%; Score 2237.5; DB 4; Length 1464;  
Best Local Similarity 60.2%; Pred. No. 1.2e-133;  
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY	1	GPP-----	-----GEPGPTGLPQPPGGRGPGSRGPGADGVAGPKGPGAGERSGPGA	48
DB	458	GPPGAGEGKRGARCEPCTCLPQPPGGRGPGSRGPGADGVAGPKGPGAGERSGPGA	517	
QY	49	GPKGSGEAGRGEGAGLPGAKLTGSPGPDGKTGTPPGAGODGRGPPGPGARGQA	108	
DB	518	GPKGSGEAGRGEGAGLPGAKLTGSPGPDGKTGTPPGAGODGRGPPGPGARGQA	577	
QY	109	GVWGPFPGKGAAGEPKAGRGVPPGAVGAGKDGAGAGAGGPPGAGPAGERGEQSPA	168	
DB	578	GVWGPFPGKGAAGEPKAGRGVPPGAVGAGKDGAGAGAGGPPGAGPAGERGEQSPA	637	
QY	169	GSFGFQGLPGAPGPEAGKPGEQVPGDLGAPGSPAGE-----	212	
DB	638	GSFGFQGLPGAPGPEAGKPGEQVPGDLGAPGSPAGE-----	697	
QY	213	-----TCLPFGGRGGRGSRGPGADGVAGPKGPGAGERSGPGAPKGPSP	258	
DB	698	GANGAPGNDGAKGDAGAPGAPGSGAGLPGKDRGDAGPKGADGSP	757	
QY	259	GE-----AGRPGEGAGLPGAK-----GLTSGSPGPPGKTGTPPGAGODGSP	300	
DB	758	KDGVGRLGTPGPPGAPGAPGDKGSGSPGAPGTGARGAPGDRGEPGPPGAPGAPGP	817	
QY	301	GPPGPGARGQ-----AGVWGPFPGKGAAGEPKAGRGVPPGPGA-----	341	



```

NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/500,811
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/187,319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)
US-09-500-811-18

Query Match 63.9%; Score 2228.5; DB 3; Length 1341;
Best Local Similarity 60.0%; Pred. No. 4.2e-133;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GEPGTLGPPGCGGSGRPPGADGVAGPKGAGRGSGPPA 48
DB 334 GPPGAGEGKRGAGEGPTGLPFGGCGGSGRPPGADGVAGPKGAGRGSGPPA 393
QY 49 GPKSGPAGRGAGLPCAKGLTSGSGSPGDPKGTGPPGAGQDGRPPGPPGARGQA 108
DB 394 GPKSGPAGRGAGLPCAKGLTSGSGSPGDPKGTGPPGAGQDGRPPGPPGARGQA 453
QY 109 GVMGPPGKGAAGEFGKAGRGVPPGAVGPAGKDGAGAGQPPGAGRGEGQGA 168
DB 454 GVMGPPGKGAAGEFGKAGRGVPPGAVGPAGKDGAGAGQPPGAGRGEGQGA 513
QY 159 GSPGQGLPGAGPGEAGKEGQVPGDLGAPGSPGAGE-----PGP--- 212
DB 514 GSPGQGLPGAGPGEAGKEGQVPGDLGAPGSPGAGE-----PGP--- 573
QY 213 -----TGLPFGCGGSGRPPGADGVAGPKGAGRGSGPPAGKSP 258
DB 574 GAGAPGNDGAKDAGAPGAGSQAGLQGWPGERGAAGLPGPKDRGDAGPKGADGSP 633
QY 259 GE-----AGRGEAGLPCAK-----GLTSGSPGPDGKTGPPGAGQDGR 300
DB 634 GKDGVRLGTGTPGPPGAPGDKGESGSPGAGTGARGAPGDRGEPGPPGAGFAGPP 693
QY 301 GPPGPPGARGO-----AGVMGPPGKGAAGEFGKAGRGVPPGPA----- 341
DB 694 GADGQFGAKGEPGDAGAKDAGPPGAPGAPGPIGNVAPGAKGARGSPGATGFP 753
QY 342 -----VGPAGKDGAGAGQPPGAGP-----AGERGQAGSGFGQGLPGPAGPPGK 393

```

RESULT 4

```

US-09-570-573-18
Sequence 18, Application US/09570573
Patent No. 6342361
GENERAL INFORMATION:
APPLICANT: Oviast, Per
APPLICANT: Bonde, Martin
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
TITLE OF INVENTION: Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/570,573
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/187,319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)

```

```

DB 754 GAAGRVGPPGSGNAGPPGPPGAGKEGKGRGTGPRGTGPRGPPGPPGPPGAGEKSP 813
QY 394 GEQGVPGDLGAPGSGPAGGEGTGLPCCGCGGSGRPPGADGVAGPKGAGRGSP 453
DB 814 GADGAPAGPPTPGQGLAGQGVVGLPQCRGERFPGLPGSGPCKGSGSASGERP 873
QY 454 GPAGP---KSPGAGRPGAGLPCAKGLTSGSGSPGDPKGTGPPGAGQDGRPPGPP 510
DB 874 GPMGPPGLAGPPGSGREGAPCAEGSPCRDGSPPGAKGDRGTGPPGPPGAGXGAGP 933
QY 511 GARGOAGVMGPPGPKGAGRGKAGRGVPPGAVGPAGKDG-----AGAQ 558
DB 934 GPAGKSGDRGTGTPAGPAGVGPAGPAGPQGRGDKGTGSGDGRGIGKRGFSLQ 993
QY 559 GPPGP-----AGPAGRGEGQAGSP---GFGGLPGPAGPGEAGKPPGQGV 603
DB 994 GPPGPPGSGGPPGSGASGPPGPPGAGKDGKGLNGLPGLPGLPGLPGRGTG 1053
QY 604 GDLGAPGSGPAG 616
DB 1054 GPPGPPGPPGPPG 1066

```



[illegible]

## RESULT 6

```

US-09-585-887-9
/ Sequence 9, Application US/09585887
/ Patent No. 6413742
/ GENERAL INFORMATION:
/ APPLICANT: Olsen, David R
/ APPLICANT: Chang, Robert
/ APPLICANT: McMullin, Hugh
/ APPLICANT: Hirzeman, Ronald A.
/ APPLICANT: Chsholm, George
/ TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
/ TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
/ TITLE OF INVENTION: CELLS
/ FILE REFERENCE: 22502030400
/ CURRENT APPLICATION NUMBER: US/09/585,887
/ CURRENT FILING DATE: 2000-05-31
/ PRIOR APPLICATION NUMBER: 09/289,578
/ PRIOR FILING DATE: 1999-04-09
/ PRIOR APPLICATION NUMBER: 60/084,828
/ PRIOR FILING DATE: 1998-05-08
/ NUMBER OF SEQ ID NOS: 11
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 9
/ LENGTH: 1461
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-585-887-9

```

Query Match 63.8%; Score 2224.5; DB 4; Length 1461;  
Best Local Similarity 59.9%; Pred. No. 8e-133;  
Matches 439; Conservative 23; Mismatches 154; Indels 117; Gaps 12;

Qy	1	GPP-----GEGPPTGLPGPGERGGSGSFGFCADGVAGPKGCPAGERGSPGPA	48
Db	455	GPFPAGBEGKKGARGEPPTGLPGPFERGGFGSGRGFGADGVAGCPKGAPAGERGSPGA	514
Qy	49	GPKGSGFBAGRPBGEAGLTPCAKGLTSGPSSGPGDGKTPGPPQPADQDRPPGPPPGPARGQA	108
Db	515	GPKGSGFGEAGRPGEAGLTPCAKGLTSGPSSGPGDGKTPGPPQPADQDRPPGPPPGPARGQA	574
Qy	109	GVWGPPFGPKGAAGEFGKAGERGVPCPGAVGPAGKGEAGAQGPPGPGPAGERGEGQGPA	168
Db	575	GVWGPPFGPKGAAGEFGKAGERGVPCPGAVGPAGKGEAGAQGPPGPGPAGERGEGQGPA	634
Qy	169	GSPGFQGLPGPAGPPGGEAGKPGEGQVCPDGLCAPGPSGAGE-----PGF----	212
Db	635	GSPGFQGLPGPAGPPGGEAGKPGEGQVCPDGLCAPGPSGAGERFFGGERGVQGPFGPAGPR	694
Qy	213	-----TGLPGPPGERGGSGSRFFPGADGVAGPKGAPAGERGSPGAPGPKSP	258
Db	695	GANGCPDNDGAKGDADAGAPGAGPSGAPGLQGMGPFERGAAGLPKPKGDRGDAGCKADGSP	754

259	QY	GE-----AGRGEAGLPGAK-----GLTSPGSGPPDGKTPGPPAGQDGRP	300
:	:	:::::	:
755	Db	GKDGVRLGTPTIGPFGPAGAPGDKGESGSPAGFTGARFAGPDGRGEPGPFPAGFAGPP	814
:	:	:::::	:
301	QY	GPFGPPGARGQ-----AGVMGPPGPKGAAGEPKGAGERGVPPGPA-----	341
:	:	:::::	:
815	Db	GADGQFCAKEGPCDDAGAKDAGCPGPPAGPAGPPGIUNVGAFAKAGARSGACPPGATGPP	874
:	:	:::::	:
342	QY	-----VGPAKDGEAGAAGPPGPAGP---AGRGEGQPAGSGPFQGLPCPAGPPGPAKXP	393
:	:	:::::	:
875	Db	GAAGRVPGPSGNAGPPCPPGPAGKEGKGSRGETGPAGRCOEYGVPPGPPPAGEKGSF	934
:	:	:::::	:
394	QY	GCQGVPDDLGAQPSGPAGEPRTLPGPPBRRGGPSRFFPGADGVAGPKGPAGERGSP	453
:	:	:::::	:
935	Db	GADGPAGAPTGTPOGIAQRWVLPGQRGERRPFLPGPSGEPKQKQPSGASGERGPP	994
:	:	:::::	:
454	QY	GPAGP---KGSPEAGRPCEAGLCAKGLTGSPSGPPDGKTGPPGPAGQDGRGPPGPP	510
:	:	:::::	:
995	Db	GFMPGPLAGPPGESREGCAPAEGSPPRGDSGFGAKGDRGETGPAGPPGAPGAPVAGPV	1054
:	:	:::::	:
511	QY	GARGQAVMGFPCKGAAGEPKKAGERGVPPGPPGAVGPAGKDE-----AGAQ	558
:	:	:::::	:
1055	Db	GPAGKSDDRGETGPAGPAGVFVGARGPAGPQGRPRDKGETGECQDRIKKHHRGSGLO	1114
:	:	:::::	:
559	QY	GPFGP-----AGPAGERGQGPAGSP--GFQGLPCPAGPPGAEKXPGSQGVP	603
:	:	:::::	:
1115	Db	GPFGPSGPEGQSASGPAGPRGPPGSAGAPGKDNLNGLPGFI GPPGPRGRTGDAGPV	1174
:	:	:::::	:
604	QY	GDLGAPGPSGPAG	616
:	:	:::::	:
1175	Db	GPFGPPGPPGPPG	1187

## RESULT 7

```

US-09-289-578-9
; Sequence 9, Application US/09289578
; Patent No. 6428978
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030450
; CURRENT APPLICATION NUMBER: US/09/289,578
; CURRENT FILING DATE: 1999-04-10
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-289-578-9

```

Query Match 63.8%; Score 2224.5; DB 4; Length 1461;  
Best Local Similarity 59.9%; Pred. No. 8e-133;  
Matches 439; Conservative 23; Mismatches 154; Indels 117; Gaps 12;

QY	1	GPP-----GEPPTGLPGPRGRCGSGRFFCADGACVAGPKGPAERGSQPA	48
DB	455	GPPGPAEGCKRGARGEPPTGLFPPGRCGSGSGFPAGDVGAFKGPAGRGSPCA	514
QY	49	GPKSGPGRAGRPCEAGLPGAKGLTGSFGSPGPKTGTPPGACQDGRPPGPPGARGQA	108
DB	515	GPKSGPGRAGRPCEAGLPGAKGLTGSFGSPGPKTGTPPGACQDGRPPGPPGARGQA	574
QY	109	GVNGFFGPKGAAGEPCKAGBERGVFPFPAVGPAGKQGEAQAQQPPGPAGPAGERGEQCPA	168





```
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 49
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-49

Query Match      62.8%; Score 2191; DB 3; Length 822;
Best Local Similarity 56.9%; Pred. No. 6.3e-131;
Matches 436; Conservative 22; Mismatches 158; Indels 150; Gaps 13;

Qy 1 GPPGEPPTGLPDPGE-----RGPGSRGFPAGDGVAGPKGAGRGSPGPA 48
Db 12 GPPGAPGQGFPGGPFGEFGSGPMGPRGPPGPKNGDDGEAGKPRGFRGPPGQ 71
Qy 49 GPKSGPCEAGRP-----GEAGLPKAGLGTGSPGSPGDKGTGPPGAGQD 93
Db 72 GARGLPTAGLPMKXHGRLGSLDKAGDAGPAGKPGSPGNGAPGQVGRGLPGR 131
Qy 94 GRPGPPPPGARGAGVGMFPKGAAGE-----PKAGERGVPPGPAVAGPKDGEA 147
Db 132 GRPGPPTAGAGNDGAVGAAGPPTGPTGTPGFPFPAVAGKAGAGPQAGRGSPQVGR 191
Qy 148 GAQPPRPAGPAGRGCGP-----AGSPGFQGLPQAGPPGAGKRGEG 192
Db 192 GEPGPPGAGAGPAGNPGADGQPGKANGAPGAGPAGPARGSPGQSGSPGPK 251
Qy 193 GVPDGLGAP-----GPSGPA-----GPPGPTGLPFPGERGGP 225
Db 252 GNSGEPAGPNKGTAKGRPGATGVQPPCPAGEEKGARGEPFGLPFPGERGGP 311
Qy 226 GSRFPAGDGVAGPKGAGRGSPGAPGKSPGSPGAGPAGPAGLPGAKGLTSPGSPGD 285
Db 312 GSRFPAGDGVAGPKGSPGSPGAPGAPGKSPGSPGAGPAGPAGLPGAKGLTSPGSPGD 371
Qy 286 KGTGPPGAGDGRPPGPPGAGCAQAGVMGFPKPAAGPAGKAGRGVPPGPAVGA 345
Db 372 KGTGPPGAGDGRPPGAGPAGCAQAGVMGFPKPAAGPAGKAGRGVPPGPAVGA 431
Qy 346 KKGAGAGQPPGAGPAGRGEGGAGSPGFGQLPQAGPPGAGKPGGQGVPGDLGAP 405
Db 432 KKGAGAGQPPGAGPAGRGEGGAGSPGFGQLPQAGPPGAGKPGGQGVPGDLGAP 491
Qy 406 GPSGAPG-----PGP-----TGLPGRPPGGRGSPGRFP 435
Db 492 GPSGARGRPFGRGVGPPGAPGPRGNNAGPNDGAKGDTGAPGAPGCGAPGLGMP 551
Qy 436 GADGVAGPKGAGRGSPGAPGKSPGCE-----AGRPGEAGLPGAKGLTSPGSP 486
Db 552 GERGAAGLPGKGRDAGPKGADGSPKDCARGLTGPIGPPGAPGAPGKAGRGSPG 611
Qy 487 GPD-----GKTGPPGAGQDGRCPGPPGPAQCG-----AGVMGFP 522
Db 612 GPTGARGAPGRGAGPPGAGPAGPAGDAGQPGKAGGPGDTGVKGADGPPGAGPAG 671
Qy 523 GPKGAAGPFGKAGRGVPPGPA-----VGPAGKDSAGAQGPPGAGP-----AGER 570
Db 672 CPGVNVGAPGPKGPRGAGPAGATGFCGAAAGRVPPGSPGNAGPPGPPGVPKGGKGR 731
Qy 571 GEQPGAGSPGQGLPQAGPAGPAGKPGEGQVPGDLAGPSPGAG 616
Db 732 GETPAGRPGEVPPGPPGAGEXGSPGADGAPGASPTGPPQGIAG 777
```

```
RESULT 10
US-08-468-996-10
; Sequence 10, Application US/08468996
; Patent No. 6645504
; GENERAL INFORMATION:
; APPLICANT: Weiner, Howard
; APPLICANT: Miller, Ariel
; APPLICANT: Zheng, Zhengli
; APPLICANT: Ahmad, Al-Sabbagh
; TITLE OF INVENTION: BYSTANDER SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION
; TITLE OF INVENTION: GLUCAGON
; FILE REFERENCE: 1010/16959-US3
; CURRENT APPLICATION NUMBER: US/08/468,996
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: US 07/843,752
; PRIOR FILING DATE: 1992-02-28
; PRIOR APPLICATION NUMBER: US 07/460,852
; PRIOR FILING DATE: 1990-02-21
; PRIOR APPLICATION NUMBER: US 07/596,936
; PRIOR FILING DATE: 1990-10-15
; PRIOR APPLICATION NUMBER: US 07/065,734
; PRIOR FILING DATE: 1987-06-24
; PRIOR APPLICATION NUMBER: US 07/454,486
; PRIOR FILING DATE: 1989-12-20
; PRIOR APPLICATION NUMBER: US 07/487,732
; PRIOR FILING DATE: 1990-03-02
; PRIOR APPLICATION NUMBER: US 07/551,632
; PRIOR FILING DATE: 1990-07-10
; PRIOR APPLICATION NUMBER: US 07/379,778
; PRIOR FILING DATE: 1989-07-14
; PRIOR APPLICATION NUMBER: US 07/607,826
; PRIOR FILING DATE: 1990-10-31
; PRIOR APPLICATION NUMBER: US 07/595,468
; PRIOR FILING DATE: 1990-10-10
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 1017
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-468-996-10

Query Match      56.0%; Score 1955; DB 4; Length 1017;
Best Local Similarity 52.5%; Pred. No. 5.4e-116;
Matches 396; Conservative 38; Mismatches 182; Indels 138; Gaps 15;

Qy 1 GPPGPPPTGLPGRGGRGPGSRGFPAGDGVAGPKPAGRGSPGAPGKSPGCEAGRP 60
Db 25 GNPGEPPGEPVSGPMGPRGPPGPPGPKPGDDGEAGPKAGRGPPGPGQARGFPPTGLP 84
Qy 61 -----GEAGLPKAGLGTSP-----GSFGP-----DGKTPPPGPA--- 90
Db 85 GVKHGRGYPGLDGAKEAGAPGVKGESGSPGNGSPGPMGRLPGERGRTGPAGAAGAR 144
Qy 91 GQDGRPPGPPGARGCAQAGVMGFP-----GPKGAGEBPKAGRGVPPGPAVGA 141
Db 145 GNDQGPAGPPGPPGVPAGPGFPAGPAGKAGAGTGARGPAGGAGPAGPAGPAGP 204
Qy 142 -----GKDGEAGAGCPGPPAGPAGRGEGQSPAGSPGFGQLPAGPAGPAGPAGP 195
Db 205 GASGNPPTGIGFAGKSGSAGAPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 264
Qy 196 GD-----LGAPGSPGAPGPGPTGLP-----GPPGRGPGSGRFPAGDVA 237
Db 265 GEQPKGEPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 324
Qy 238 GPKGAPGREGSPGAPGKSPGCEAGRPGEAGLPGAKGLTGTSPGSPGPDGKTPGPPAGQD 297
Db 325 GPKGAPGREGSPGAPGKSPGCEAGRPGEAGLPGAKGLTGTSPGSPGPDGKTPGPPAGQD 384
Qy 298 GRPFPFPFGARQAGVMGFPQPKGAAGBPKAGRGVPPGPAVAGPAGKAGEAGAGQGP 357
```

```

385 GRPFPFGGAGQGQGVWGFPPKGANGPCKAGKGLGAPGURLGLPKDGETGAEGPP 444
358 GPAGPAGERGEGQAGSPGFGQLGPGAPPGEGEAGKPGEGVPGDGLGAPGSPGAGE- 413
445 GPAGPAGERGEGQAGSPGFGQLGPGPPGEGEAGKPGDQGVPGGAGAGLVGPRGERGPP 504
414 -----PGPTGLP-----GPPGCGGSGRFFPGADGVAGPKGA 447
505 GERGSFGAGQLGPPGLFTGTGTPGKASGAPGPPGAQGPGLQGMPPGERGAAGIAGPK 564
448 GERGSFGPAGPKGSFGE-----AGRPGEAGLPGAKGLTGPSPSPQPD 489
565 GDRGVGKEGPEGAPGKDGARGLTGTPPGPAGANGKEGVEVPPGAGSAGARGAPGER 624
490 GKTGPPGAGQDGRGPPGPGAR-----GQGVWGFPPKKAAGEP-----GKA 534
625 GETGPPGAGFAGPPGADQGVGKAGEGEGAGKGDAGAPGQPSGAPGQGTGVTGPK 684
535 GERGVGPPPGA-----VGPAGKDGEGAGAGQPPGPA---GPAGERGEGQAGSPGPFQ 582
685 GARGAQGPPGATPPGAAGRVGPPGSGNGNPPGPPGPPGSKDGKARGDSGPPGRAGER 744
583 GLPGPAGPPGAGKPGEGQGVPGDGLGAPGSGPAG 616
745 GLQGAGPPGKGBGPDGPGSAGPPGQGLAG 778

RESULT 11
US-08-931-820-3
; Sequence 3, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type II
; US-08-931-820-3

```

QY	91	GDGRPPGPPGGARGAQGVVGF-----GPKGAAAGEPGKAGRGVPPGPAVGPA	141
Db	164	GNDGQFGPAGPPGVGVPAGPFPGFAPGAKGEAGPTGARGPGEAQOFPREPTPGSPGA	223
QY	142	-----GXDGABAGACGPPGAPGAGEREGCPAGSDPFOGLPDPAGFPPEAGKPGEQGV	195
Db	224	GASGNPFTDGLIPCAKGSAGAFGTAGAPFFGPRGPDPOCATGTLPGKQTCKRGIAGEPK	283
QY	196	GD--LGAAPSGSAGBPQPTGLP-----GPPERGGPSRGFFPGADGVA	237
Db	284	GEQCPKGFPGPAGPQGAPFPAGEBECKTARGEPGCVGPICPPGERGAPNRRGFPQDGLA	343
QY	238	GPKPAGGERGSPGAPGPKSGPEAGRPEAGLTGSKXLGTSKSGPCPDGKTTPCGPAGOD	297
Db	344	GPKNAPGERGSLUAGPKGANGDFGRPGEPLPGARGLTGRLPDADGPKQKVGPSGAPGED	403
QY	298	CRPPPPPARGAQGVWGMFPFKGAAGEFGKAGERGVFPFCAGVGPAGKDGEAGAQQPP	357
Db	404	CRPGPPGQARGQPVMGFPFKGANGEBPKAGEKGLPGAELRGLPGKDCETCAEGPP	463
QY	358	GPAGPAGERGQAGSPGFOGLPGPAGPPEACEACKSEOGVPGDLGAPCPSPAGE----	413
Db	464	GPAGPAGERGQAGPSPGFOGLPGPPPPGEGCKPGDQGVPEBAGAPLVGPRGERGFP	523
QY	414	-----PGPTGLP-----GPPERGGPSRGFPAGADGVAGPKGA	447
Db	524	GERGSPNAQLGPRGLIPTCTGTDKGASGPAPGPAAGPPFLQGMFERGAAAGIAPK	583
QY	448	BERGS-----PGPAGPKGSPCEAGRPEAEGLPGAKGLTGP	483
Db	584	GDREVKEKPEGAFPXKDGGRGLTGTTPGPAGANGEKGVPPGCPAGSAGAGAPGER	643
QY	484	GSPGDGKTGPPGAGODGRGPPGPAEGQGVWGMFPFKGAAGEP-----GKA	534
Db	644	GETPGTGTGIAGPPGADGQ-GAKGQGBAGQKGDAGAPGQPSGAPQPGPTVTGPK	703
QY	535	GERGVPPGPA-----VGPAGKDGAGAACGPPGPA---GPAGERGEOGPAQSPOFQ	582
Db	704	GARGAQPPGATGPPGAAGRVGPPGNGNPDPGPPGSKDKGPKGARDSGPPGRAGEP	763
QY	593	CLPFPAGPEAGKPEQGVPLGARGSPGAP	616
Db	764	GLOPPAGPPEKPEBDDGSPGAEPPGQGLAG	797
 RESULT 12 US-08-963-825-20 ; Sequence 20, Application US/08963825 ; Patent No. 6110589 ; GENERAL INFORMATION: ; APPLICANT: Qvist, Per ; APPLICANT: Bonde, Martin ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments ; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence ; TITLE OF INVENTION: Disorders Associated with the Metabolism of ; NUMBER OF SEQUENCES: 21 ; CORRESPONDENCE ADDRESS: ; ADDRESSEE: Darby & Darby PC ; STREET: 805 Third Avenue ; CITY: New York ; STATE: New York ; COUNTRY: USA ; ZIP: 10022 ; COMPUTER READABLE FORM: ; MEDIUM TYPE: Floppy disk ; COMPUTER: IBM PC compatible ; OPERATING SYSTEM: PC-DOS/MS-DOS ; SOFTWARE: PatentIn Release #1.0, Version #1.25 ; CURRENT APPLICATION NUMBER: US/08/963,825 ; FILING DATE: ; CLASSIFICATION: 436			

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/187,319  
FILING DATE: 21-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: GOSORIS, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1418 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN -ALPHA 1 (II)  
US-08-963-825-20

Query Match 55.7%; Score 1943; DB 3; Length 1418;  
Best Local Similarity 52.1%; Pred. No. 4e-115;  
Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGEPGTLGPPGGRGSGRGGFPGADGVAGPKGAGRGSPGAGKSPGAGRP 60  
DB 156 GNPGEPEGVSGPMGRGPPGPKGDDGEGAGKFGKAGRGPPGPGQARGFPGLP 215  
QY 61 -----GEAGLPGAKGLTGSP-----GSPGP-----DGKTGPPGPA--- 90  
DB 216 GVKHRYGVLGDAKAGAGAPGVKSGSGFENGSPGPMGRGLPGRGRTGAGAGAR 275  
QY 91 GQDGRPGPPGPPGARGAGVGMFP-----GPKGAAGEPGKAGRGVPGPPGAGPA 141  
DB 276 GNDQGPAGPPGPPGVGAGGPPGPGACAKGEAGTGAERGAGGPPGPPGPPGPA 335  
QY 142 -----GKDEAGAGPPGAGPAGRGEGGAGSPGFGPLGPPAGPPGAGKGGQVP 195  
DB 336 GASGNPGTDGIPGAKGSAGAPGAGAPGPPGPPGPPGPPGPPGPPGPPGPPG 395  
QY 196 GD-----LGAPGSPGAGPPGPTGLP-----GPPRGGPPGSRGPPGADGVA 237  
DB 396 GEQPKGEFPGAGPQAGPAGPAGESGKRGARPGVGPPIGPPGERGAPGNRPPGQDGLA 455  
QY 238 GPKGAGERGPPGAGPKGSPGEGRPGEGAGLPGAKGLTGSPGSPGPPGPPGAGQD 297  
DB 456 GPKGAGERGPPGAGPKGANGDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 515  
QY 298 GRPGRPPGARGAGVGMFPGKGAAGEPGKAGRGVPGPPGAGVPGAGKDGAGAGOPP 357  
DB 516 GRPGRPPGARGAGVGMFPGKGAAGEPGKAGRGVPGPPGAGVPGAGKDGAGAGOPP 575  
QY 358 GPAGPAGERGPPGAGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 413  
DB 576 GPAGPAGERGPPGAGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 635  
QY 414 -----PGTGLP-----GPPRGGPPGSRGPPGADGVAGPKGPA 447  
DB 636 GERGSPGAGLQGRGRLGPTGTGDPKGAAGPAGPPGAGPPGPPGPPGPPGPPGPPG 695  
QY 448 GERGS-----PGPAGPKSPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 483  
DB 696 GDRGIVGKEGEGAGKDGGRGLTGPIGPPGAGANGKGVGPPGPPGAGSAGAGAPGER 755  
QY 484 GSPGPDGKTGPPGAGQGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 534  
DB 756 GETGPPGTSGIAGPPGADGQPCAKGEQGEAGCKDAGAPGPPGPPGPPGPPGPPGPPG 815  
QY 535 GERGVPPGPPG-----VGPAGKDCGAGAGPPGPA-----GPAGRGEGPPGSPGFQ 582

DB 816 GARGAQPPGATGPPGAGRGVPGSGNGNPPGPPGPPGPPGPPGPPGPPGPPGPPG 875  
QY 583 GLPGFAGPPGAGKPGEGQGVGPDGLGAPGPPGPPG 616  
DB 876 GLQGFAGPPGEGKPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 909

RESULT 13  
US-09-010-999-1  
Sequence 1, Application US/09010999  
Patent No. 6132976  
GENERAL INFORMATION:  
APPLICANT: Poole, Anthony R.  
APPLICANT: Hollander, Anthony P.  
APPLICANT: Ballinghurst, R. C.  
TITLE OF INVENTION: IMMUNOASSAYS FOR THE MEASUREMENT OF  
TITLE OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/010.999  
FILING DATE: 22-JAN-1998  
CLASSIFICATION: 4335  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/448,501  
FILING DATE: 17-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/984,123  
FILING DATE: 04-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 032931/0212  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1418 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Human Type II Collagen  
US-09-010-999-1

Query Match 55.7%; Score 1943; DB 3; Length 1418;  
Best Local Similarity 52.1%; Pred. No. 4e-115;  
Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGEPGTLGPPGGRGSGRGGFPGADGVAGPKGAGRGSPGAGKSPGAGRP 60  
DB 156 GNPGEPEGVSGPMGRGPPGPKGDDGEGAGKFGKAGRGPPGPGQARGFPGLP 215  
QY 61 -----GEAGLPGAKGLTGSP-----GSPGP-----DGKTGPPGPA--- 90  
DB 216 GVKHRYGVLGDAKAGAGAPGVKSGSGFENGSPGPMGRGLPGRGRTGAGAGAR 275  
QY 91 GQDGRPGPPGPPGARGAGVGMFP-----GPKGAAGEPGKAGRGVPGPPGAGPA 141



QY 583 GLFGPAGPPGAGKPGEGVQGVGDLGAPGSPGAG 616  
 Db 876 GLQGPAGPGEKPGDGDGSGAEGPPGQGLAG 909

RESULT 15  
 US-09-570-573-20  
 ; Sequence 20, Application US/09570573  
 ; Patent No. 6342361  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Qvist, Per  
 ; APPLICANT: Bonde, Martin  
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
 ; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Darby & Darby PC  
 ; STREET: 805 Third Avenue  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10022  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/570,573  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/187,319  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Gogoris, Adda C  
 ; REGISTRATION NUMBER: 29,714  
 ; REFERENCE/DOCKET NUMBER: 4305/08701  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-527-7700  
 ; TELEFAX: 212-753-6237  
 ; TELEX: 236687  
 ; INFORMATION FOR SEQ ID NO: 20:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1418 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Homo sapiens  
 ; IMMEDIATE SOURCE:  
 ; CLONE: COLLAGEN -ALPHA 1 (II)

Search completed: October 18, 2004, 13:43:55  
 Job time : 29.4569 secs

Query Match 55.7%; Score 1943; DB 3; Length 1418;  
 Best Local Similarity 52.1%; Pred. No. 4e-115;  
 Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGEPGTGLPGRPGERGSGRFGPCADGVAGKPGAGRGSPGAPGKSGFGEAGRP 60  
 Db 156 GNPGEFPGVSGPMGPRGPPGPKPGDDGEGAKPGKAGRGPPGQARGPFGTGLP 215

QY 61 -----GEAGLFGAKGLTGPSPG-----GSPGP-----DGKGTGPPGPA--- 90  
 Db 216 GVKGHRGYPLDGAKEAGAGVKGESGSPGPMGRGLPGERGRTGPAGAAGAR 275

QY 91 GQDGRPGPPGPPGARGQGVGVGFP-----GPKGAAGPPGKAGRGVPPGPPGAVGPA 141  
 Db 276 GNDQGPAGPPGPPGVPAGGPGFPCAPGAKGAGTCAARGPEGAQGRGPFGTGSPGPA 335

QY 142 -----GKDGAGAGGPPGPPGAGRGEGEQGPGSPGQGLPFGPPGPPGAGKPGEQGVP 195

Db 336 GASGNPVDGTPGAKGSAGAPGAGPFGPFRPPDPQATGLGLGKQTKGFIAGFK 395  
 QY 196 GD---LGAPGSPGAPGFGTGLP-----GPPGERGPPGSRGFFPADGVA 237  
 Db 396 GEQPKGEPGAPGQAGPAGGEGKRGARGEPGVGPIGPPGERGAPGNRGFFQDGLA 455  
 QY 238 GPKGPAGRGSPGAPGKSGPGEAGRGKGLTGSPPSGPDGKTGPPGAGQD 297  
 Db 456 GPKGAPGERGSPGLAGPKXGANGDPGRGEPGLRGARLTGPPGAGKQKVGPPGAPGED 515  
 QY 298 GRPGRPGLPARGQAGVGMGFPKPKGAGEPKAGRGVPPGPPGAVGPPAGKDEAGAAQPP 357  
 Db 516 GRPGRPGLQARGGPPGVMGFPKPKGANGEPCKAGEKGLGAPGLRGLPKDGETGAEPP 575  
 QY 358 GPAGPAGRGSGPAGSPGFGPLPQAPGPPGKAGEKPGEQGVGDLGAPGSPGAGE---- 413  
 Db 576 GPAGPAGRGSGPAGSPGFGPLPQAPGPPGKAGEKPGEQGVGDLGAPGSPGAGE---- 635  
 QY 414 -----PGTGLP-----GPPGERGPPGSRGFFGADGVAGPKGPA 447  
 Db 636 GERGSPGAGLQGRGLPGTGTGDPKASGAPGPPGAGQPPGLQGMFGERGAAGIAGPK 695  
 QY 448 GERGS-----PPGPKSGPGEAGRPGEAGLFGAKGLTGPSP 483  
 Db 696 GDRGDVGEKPEGAPCKDGRGLTGPIGPPGAPAGANGKEGVGPPGPPGASAGARGAPGER 755  
 QY 484 SSPGPDGKTGPPGAPGQDGRGPPGPPGARGQAGVGMGFPKPKGAGEP-----GKA 534  
 Db 756 GETGPPGTSGTAGPPGADGQCAKGEQGEAGQKGDAGAPGPPGPPGAPGPPGTPGTGPK 815  
 QY 535 GERGVGPPGPA-----VGPAGKDGEGAAGQPPGPA---GPAGERGEQGPAGSPGFQ 582  
 Db 816 GARGAQQPPGATGFFGAAAGRVGPPGSGNNGPPGPPGSPGKDGKARGDGGPPGRAGEP 875

QY 583 GLPFPAGPPGKAGEKGEQGVGPDILGAPGSPGAG 616  
 Db 876 GLQGPAGPGEKPGDGDGSGAEGPPGQGLAG 909

*This Page Blank (uspio)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:42:06 ; Search time 79.3776 Seconds  
(without alignments)  
2513.149 Million cell updates/sec

Title: US-10-658-989A-2  
 Perfect score: 3488  
 Sequence: 1 GPPGEPGPTGLPDPGERG.....GEQVPGDLGAPSPGPAG 617

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1360919 seqs, 323318874 residues  
Total number of hits satisfying chosen parameters: 1360919

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```

Database : Published Applications AA: *
1: /cgn2_6/ptodata/2/pubpaa/US07 PUBCOMB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US06 PUBCOMB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/US06 PUBCOMB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US07 NEW PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS PUBCOMB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/US08 NEW PUB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/US08 PUBCOMB.pcp.*
9: /cgn2_6/ptodata/2/pubpaa/US09A PUBCOMB.pcp.*
10: /cgn2_6/ptodata/2/pubpaa/US09B PUBCOMB.pcp.*
11: /cgn2_6/ptodata/2/pubpaa/US09C PUBCOMB.pcp.*
12: /cgn2_6/ptodata/2/pubpaa/US09C NEW PUB.pcp.*
13: /cgn2_6/ptodata/2/pubpaa/US10A PUBCOMB.pcp.*
14: /cgn2_6/ptodata/2/pubpaa/US10B PUBCOMB.pcp.*
15: /cgn2_6/ptodata/2/pubpaa/US10C PUBCOMB.pcp.*
16: /cgn2_6/ptodata/2/pubpaa/US10D PUBCOMB.pcp.*
17: /cgn2_6/ptodata/2/pubpaa/US10 NEW PUB.pcp.*
18: /cgn2_6/ptodata/2/pubpaa/US60 NEW PUB.pcp.*
19: /cgn2_6/ptodata/2/pubpaa/US60 PUBCOMB.pcp.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	2237.5	64.1	1057	15	US-10-104-889-20		Sequence 20, Appl
2	2237.5	64.1	1107	15	US-10-104-889-11		Sequence 11, Appl
3	2237.5	64.1	1171	15	US-10-104-889-8		Sequence 8, Appl
4	2237.5	64.1	1368	15	US-10-104-889-10		Sequence 10, Appl
5	2237.5	64.1	1464	14	US-10-216-705-21		Sequence 21, Appl
6	2237.5	64.1	1464	16	US-10-357-851-1		Sequence 1, Appl
7	2237.5	64.1	1464	16	US-10-358-024-1		Sequence 1, Appl
8	2237.5	64.1	1464	17	US-10-788-792-150		Sequence 150, App
9	2234	64.0	1449	15	US-10-402-089-8		Sequence 8, Appl
10	2234	64.0	1449	15	US-10-402-0728-8		Sequence 8, Appl
11	2233.5	64.0	1461	16	US-10-468-091-25		Sequence 25, Appl
12	2233.5	64.0	1464	14	US-10-591-265-243		Sequence 243, App
13	2231.5	64.0	1341	15	US-10-104-889-16		Sequence 16, Appl
14	2228.5	63.9	1341	14	US-10-058-124-18		Sequence 18, Appl

15	2227.5	63.9	1464	10	US-09-918-715-261	Sequence 261, Appl
16	2227.5	63.9	1464	14	US-10-060-036-159	Sequence 159, Appl
17	2227.5	63.9	1464	14	US-10-171-311-36	Sequence 36, Appl
18	2227.5	63.9	1464	14	US-10-149-352-2	Sequence 2, Appl
19	2227.5	63.9	1464	14	US-10-177-393-65	Sequence 65, Appl
20	2227.5	63.9	1464	14	US-10-301-832-28	Sequence 28, Appl
21	2227.5	63.9	1464	16	US-10-734-864-79	Sequence 79, Appl
22	2227.5	63.8	1463	15	US-10-402-089-2	Sequence 2, Appl
23	2225.5	63.8	1463	15	US-10-402-072A-2	Sequence 6, Appl
24	2225.5	63.8	1463	15	US-10-104-889-6	Sequence 6, Appl
25	2225.5	63.7	1463	15	US-10-104-889-6	Sequence 26, Appl
26	2205.5	63.2	1453	16	US-10-468-091-26	Sequence 49, Appl
27	2191	62.8	1452	14	US-10-342-331-49	Sequence 49, Appl
28	1946	55.8	1014	14	US-10-639-286-10	Sequence 10, Appl
29	1943	55.7	1418	14	US-10-194-441A-1	Sequence 1, Appl
30	1943	55.7	1418	16	US-10-058-124-20	Sequence 20, Appl
31	1932	55.4	1014	14	US-10-468-091-5	Sequence 5, Appl
32	1932	55.4	1014	14	US-10-194-441A-48	Sequence 48, Appl
33	1932	55.4	1487	16	US-10-468-091-6	Sequence 6, Appl
34	1919	55.0	714	14	US-10-233-895-44	Sequence 44, Appl
35	1919	55.0	714	14	US-10-231-581-44	Sequence 44, Appl
36	1892.5	54.3	662	14	US-10-326-508A-15	Sequence 15, Appl
37	1885	54.0	505	14	US-10-232-175-37	Sequence 37, Appl
38	1882.5	54.0	595	14	US-10-232-175-17	Sequence 17, Appl
39	1882.5	54.0	595	14	US-10-342-331-48	Sequence 48, Appl
40	1875	53.8	1466	15	US-10-342-331-50	Sequence 50, Appl
41	1875	53.8	1466	15	US-10-402-089-12	Sequence 12, Appl
42	1862	53.4	1466	15	US-10-402-072A-12	Sequence 4, Appl
43	1862	53.4	1466	15	US-10-403-089-6	Sequence 6, Appl
44	1862	53.4	1466	15	US-10-403-089-6	Sequence 4, Appl
45	1862	53.4	1466	15	US-10-402-072A-6	Sequence 6, Appl

## ALIGNMENTS

RESULT 1  
US-10-104-899-20  
; Sequence 20, Application US/10104899  
; Publication No. US20040086961A1  
; GENERAL INFORMATION:  
; APPLICANT: GRUSKIN, ELLIOT A.  
; BUECHTER, DOUGLAS  
; BROKAW, JANE  
; ZHANG, GUANGHUI  
; PAOLELLA, DAVID  
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DILWORTH & BARRESE  
; STREET: 333 EARLE OVINGTON BOULEVARD  
; CITY: UNIONDALE  
; STATE: NY  
; COUNTRY: U.S.A.  
; ZIP: 11553  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/104,889  
; FILING DATE: 22-Mar-2002  
; CLASSIFICATION: <Unknown>  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US/09/169,768  
; FILING DATE: 09-OCT-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEEN, JEFFREY S  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (516) 228-8484  
; TELEFAX: (516) 228-8516  
; INFORMATION FOR SEQ ID NO: 20:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID N
US-10-104-889-20

```

Query Match 64.1%; Score 2237.5; DB 15; Length 1057;  
Best Local Similarity 60.2%; Pred. No. 1.9e-104;  
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

```
QY      1 GPP-----GGPGTGLPPGREGGSGSFFGADGVAGPKPGACERGSQCPA 48
DB      297 GPPGAPAGEGKRGARGGPGFTGLPPGREGGSGSGFFGADGVAGPKPGACERGSQCPA 356
QY      49 GPKSGFGAGRPGEAGLPGAKGLTSGSGSPGDKTTPPGPAGQDQRPGPPGPGCARQA 108
DB      357 GPKSGFGAGRGEAGLPGAKGLTSGSGSPGDKTTPPGPAGQDQRPGPPGPGCARQA 415
QY      109 GYMGPPGPKGAAGPFGKAGRGVFPPPGAVGPKDGEAGAGQPPGAPGAGERGEOGPA 168
DB      417 GYMGPPGPKGAAGPFGKAGRGVFPPPGAVGPKDGEAGAGQPPGAPGAGERGEOGPA 476
QY      169 GSPGFGQLPPGAPGPPGAGKPGEGVPGDLGAPGSPSPAGE-----GPP--- 212
DB      477 GSPGFGQLPPGAPGPPGAGKPGEGVPGDLGAPGSPGARGRGFPGERGVGPPGAPGR 536
QY      213 -----TGLPGPPGERGPGSRGPFPGADGVAGPKPGAGERSGPCAGPGXSP 258
DB      537 GANGAPGNDGAKGAGAPGAFGSGQAFGLQCMFGERGAAGLPGPKGDRGDAGPKADGSP 596
QY      259 GE-----AGRPGAGLPGAK-----GLTSGSPSGPDPCKTTPPGPAGQDGRP 300
DB      597 KDGVRGLGTIGIPGPPAGAPGDKBSGSPGAGPTGARFAGPDRGEFGPPGPGAGPP 656
QY      301 GPPGPPGARGQ-----AGVMGPPGPKGAAGEPKGAGERGVPPGPCA----- 341
DB      657 GADGQPAKGPEDGAKGDAGPPGAPGPPGPIGNVGAFCAGKAGSGAPGATGPP 716
QY      342 -----VGPAGKDGAGAQGPPGPAGP---AGRGGSGGPAAGSPGFOGLPGPAGPGEAGKP 393
DB      717 GAARGVPPGPPSGNAGPPGPPGAGKEGKGKGRGETPGAGRPGEVGPVPPGPGPAGEXGSP 776
QY      394 GEQGVPGDLGAPGSPGAPGPPGTGLPGPPGERGGSGRGGPPGADGVAGPKGAPAGERGSP 453
DB      777 GADGPACAPGTPGQGIAGQGVVGLPGQRGERGPPGLPGSPGBPKGKPGSGAERGGPP 836
QY      454 GPAGP---KGSPPAGRPGEAGLPGAKGLTSGSPGPGDKTTPPGPAGQDGRGPPGPP 510
DB      837 GPMGPPGLAGPPGESCREGAPGAEGSPGRDGSFCAKGRGETGPGAGPPGAPGAPGAPGV 896
QY      511 GARGQAGVMGPPGPKGAAGEPKKAGRGVFPPPGAVGAGPKDGE-----AGAQ 558
DB      897 GPAGKSGDRGETGPGAPGAPVPGAGARGPAGPQGRGDKGTGQGGDRGIKGHRGFSGLQ 956
QY      559 GPPGP-----AGPAGERGQGPAGSP---GFQGLPGPAGPPGAGKPGEGOV 603
DB      957 GPPGPGSPGEGQGPSGASGAPGRGPPGSAAGAFKDXGLNGLPGTIGTPGPRGRTGDAGPV 1016
QY      604 GDLGAPGSPGAP 616
DB      1017 GPPGPPGPPGPPG 1029
```

BROKAW, JANE  
ZHANG, GUANGHUI  
PAOLELLA, DAVID  
TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: DILWORTH & BARRESE  
STREET: 333 EARLE OVINGTON BOULEVARD  
CITY: UNIONDALE  
STATE: NY  
COUNTRY: U.S.A.  
ZIP: 11553  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/104,889  
FILING DATE: 22-Mar-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/169,768  
FILING DATE: 09-OCT-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: STEEN, JEFFREY S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (516) 228-8484  
TELEFAX: (516) 228-8516  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1107 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 11:  
US-10-104-889-11

Query Match	64.1%	Score 2237.5	DB 15	Length 1107	
Best Local Similarity	60.2%	Pred. No. 2e-104			
Matches 441	Conservative 23	Mismatches 152	Indels 117	Gaps 13	
QY	1	GPP-----	-----GPGPGTGLPGPPGRRGGPGSRGPGADGVAGPKPAGERGSGCPGA	48	
DB	297	GPPPAGEEGKRGARGEPGTGLPGPPGRRGGPGSRGPGCADGVAGPKPAGERGSGCPGA	356		
QY	49	GPKSGPEAGRPBAGLPGAKGLTGSFGSPDDGKTTGPPGPAQDGRGPPGPPGARGQA	108		
DB	357	GPKSGPEAGRPBAGLPGAKGLTGSFGSPDDGKTTGPPGPAQDGRGPPGPPGARGQA	416		
QY	109	GVMGFPFPKAAAGBPGKAGRGVPGPPGAVGPGAGKDGAGACGPPGPPGAPAGERGEOCPA	168		
DB	417	GVMGFPFPKAAAGBPGKAGRGVPGPPGAVGPGAGKDGAGACGPPGPPGAPAGERGEOCPA	476		
QY	169	GSPFGQLPGPAGPPBAGKPGEBQGVPGDLGAPGSGPAGE-----	PGP-----	212	
DB	477	GSPFGQLPGPAGPPBAGKPGEBQGVPGDLGAPGSGPAGE-----	PGP-----	536	
QY	213	-----TGLPPPGRRGGPGSRGPPFGADGVAGPKPAGERGSGPPGAPKGP	258		
DB	537	GANGAPNDGAKGDAGAPGAPGSGAPFLQGMFGERGAAGLPGPKGDRGDAGPKGADGSP	596		
QY	259	GE-----	AGRPGEAGLPGAK-----	GLTSGPSGPGDKTTPGPPAGQDGRP	300
DB	597	GKDGVRLGTPIGPPGAPAGDCKBESGSGPAGTGTARGAPGDRGEPGPPGPPAGPAGPP	656		
QY	301	GPPGPPGARGQ-----	-----AGVMGFPFGKAAAGBPGKAGRGVPGPPGA-----	341	
DB	657	GADQQGAKGPEPDCAKAGDAGPPGPPGAPGAPPGPIGNVGAPFAGKARGAGSGPPGATGFP	716		
QY	342	-----	VGPAGKDGAEAGQAQPPGPPAGP-----	AGBRGQGPAGSPFGQLPGPAGPPGCEAGKP	393

RESULT 2  
US-10-104-889-11  
; Sequence 11, Application US/101:04889  
; Publication No. US20040086961A1  
; GENERAL INFORMATION:  
; APPLICANT: GRUSKIN, ELLIOT A.  
; BUECHTER, DOUGLAS



[illegible]

RESULT 3  
 US-10-104-889-8  
 ; Sequence 8, Application US/10104889  
 ; Publication No. US20040086961A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GRUSKIN, ELLIOT A.  
 ; BUECHTER, DOUGLAS  
 ; BROKAW, JANE  
 ; ZHANG, GUANGHUI  
 ; PAOLELLA, DAVID  
 ; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES  
 ; NUMBER OF SEQUENCES: 50  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: DILWORTH & BARRESE  
 ; STREET: 333 EARLE OVINGTON BOULEVARD  
 ; CITY: UNIONDALE  
 ; STATE: NY  
 ; COUNTRY: U.S.A.  
 ; ZIP: 11553  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/104,889  
 ; FILING DATE: 22-Mar-2002  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/169,768  
 ; FILING DATE: 09-OCT-1998  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: STEEN, JEFFREY S  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (516) 228-8484  
 ; TELEFAX: (516) 228-8516  
 ; INFORMATION FOR SEQ ID NO: 8:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1171 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: peptide  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
 ; PS-10-104-889-8

Query Match	64.1%;	Score 2237.5;	DB 15;	Length 1171;
Best Local Similarity	60.2%;	Pred. No. 2.1e-104;		
Matches 441;	Conservative 23;	Mismatches 152;	Indels 117;	Gaps 12;

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/104,889  
 FILING DATE: 22-Mar-2002  
 CLASSIFICATION: <Unknown>  
 PRIORITY INFORMATION:  
 APPLICATION NUMBER: US/09/169,768  
 FILING DATE: 09-OCT-1998  
 ATTORNEY/AGENT INFORMATION:  
 NAME: STEEN, JEFFREY S  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (516) 228-8484  
 TELEFAX: (516) 228-8516  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1388 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-10-104-889-10

Query Match 64.1%; Score 2237.5; DB 15; Length 1388;  
 Best Local Similarity 60.2%; Pred. No. 2.3e-104;  
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GPPGTGLPPPPGRRGGSGRFPAGDGVAGKGPAGRGSPGA 48  
 DB 297 GPPGAGEGKRGARGPPTGLPPPPGRRGGSGRFPAGDGVAGKGPAGRGSPGA 356  
 QY 49 GPKSGEAGRPAGLPAGKGLTSGSPGPKTTPPGAGDGRPPGPPGARGQA 108  
 DB 357 GPKSGEAGRPAGLPAGKGLTSGSPGPKTTPPGAGDGRPPGPPGARGQA 416  
 QY 109 GVMGFPKGAAGPFGKAGRGVPPGCAVGPAGKDGAGAGAGPPGPPGARGQA 168  
 DB 417 GVMGFPKGAAGPFGKAGRGVPPGCAVGPAGKDGAGAGAGPPGPPGARGQA 476  
 QY 169 GSPGFQGLPPAGPPGAGPGEQGVGDLGAPGSPGAGE-----PGP---- 212  
 DB 477 GSPGFQGLPPAGPPGAGPGEQGVGDLGAPGSPGAGE-----PGP---- 536  
 QY 213 -----TGLPFPGERGGSGRFPAGDGVAGKGPAGRGSPGAPGKSP 258  
 DB 537 GAGAPGNDGAKGADAGAPGAPGSGAGPGLQMPGGERGAAGLPKPKDGRDAGPKGADGSP 596  
 QY 259 GE-----AGEPGEAGLPKAK-----GLTSGSPGPDGKTTPPGAGDGRP 300  
 DB 597 GKDGVRGLTGPFGPPGAGAPGDKGSGSPGAPGPTGARGAPGDRGEPGPPGAPGAPG 656  
 QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGPFGKAGRGVPPGPPGA---- 341  
 DB 657 GADGQPKAGEPDGAKGADAGPPGAPGAPGPIGNVGAFCAGKARGSPGATGPP 716  
 QY 342 -----VGPAGKDEAGAGQPPGAPG-----AGERGEGPAGSPGFGQLPPGAPGGEAGKP 393  
 DB 717 GAAGRVPFPSPGNAGPPGPPGAGKEGKGPRGETGAPRGCEVGPVPPGPPGAPGKSP 776  
 QY 394 GEQGVFDLCAAPSPGAPGPPGPTGLPPPPGRRGGSGRFPAGDGVAGKGPAGRGSP 453  
 DB 777 GADGAPAGPTFPQGIAGQGVVGLPQGRGRRFPGLPSPGEGPKQFSGASGERGPP 836  
 QY 454 GPAGP---KSGPGEAGRPGEAGLPAGKGLTSGSPGPDGKTTPPGAGDGRPPGPP 510  
 DB 837 GPMGFPGLAGPPGESREGAPGAGSGPGRDGSFGAKGDRGETGAPGPPGAPGAPGVP 896  
 QY 511 GARGQAGVMGFPKGAAGPFGKAGRGVPPGCAVGPAGKDG-----AGAG 558  
 DB 897 GPAGKSGDRGETGAPGAPGVPAGARGPAGPQGRGDRGETGEOGDRGIKHGRFSG 956

QY 559 GPPGP-----AGPAGERGEQGPAGSP-----GFQGLPGPAGPFGGAGKPGQGV 603  
 DB 957 GPPGPPGSGEQGSGAGSPAGPRGPPGSGAGAPGKXGLNGLPGTGPGRGRTGDAGPV 1016  
 QY 604 GDLGAPGSGPPAG 616  
 DB 1017 GPPGPPGPPGPPG 1029

# RESULT 5

US-10-216-705-21  
 ; Sequence 21, Application US/10216705  
 ; Publication No. US20030096973A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Meristem Therapeutics, S.A.  
 ; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me  
 ; FILE OF INVENTION: Obtaining Such and Their Uses  
 ; FILE REFERENCE: 1149-3 DIV  
 ; CURRENT APPLICATION NUMBER: US/10/216,705  
 ; PRIOR FILING DATE: 2002-08-09  
 ; PRIOR APPLICATION NUMBER: US 09/331,347  
 ; PRIOR FILING DATE: 1999-08-17  
 ; NUMBER OF SEQ ID NOS: 22  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 21  
 ; LENGTH: 1464  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-216-705-21

Query Match 64.1%; Score 2237.5; DB 14; Length 1464;  
 Best Local Similarity 60.2%; Pred. No. 2.4e-104;  
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GPPGTGLPPPPGRRGGSGRFPAGDGVAGKGPAGRGSPGA 48  
 DB 458 GPPGAGEGKRGARGPPTGLPPPPGRRGGSGRFPAGDGVAGKGPAGRGSPGA 517  
 QY 49 GPKSGEAGRPAGLPAGKGLTSGSPGPKTTPPGAGDGRPPGPPGARGQA 108  
 DB 518 GPKSGEAGRPAGLPAGKGLTSGSPGPKTTPPGAGDGRPPGPPGARGQA 577  
 QY 109 GVMGFPKGAAGPFGKAGRGVPPGCAVGPAGKDGAGAGAGPPGPPGARGQA 168  
 DB 578 GVMGFPKGAAGPFGKAGRGVPPGCAVGPAGKDGAGAGAGPPGPPGARGQA 637  
 QY 169 GSPGFQGLPPAGPPGAGPGEQGVGDLGAPGSPGAGE-----PGP---- 212  
 DB 638 GSPGFQGLPPAGPPGAGPGEQGVGDLGAPGSPGAGE-----PGP---- 697  
 QY 213 -----TGLPFPGERGGSGRFPAGDGVAGKGPAGRGSPGAPGKSP 258  
 DB 698 GAGAPGNDGAKGADAGAPGAPGSGAGPGLQMPGGERGAAGLPKPKDGRDAGPKGADGSP 757  
 QY 259 GE-----AGEPGEAGLPKAK-----GLTSGSPGPDGKTTPPGAGDGRP 300  
 DB 758 GKDGVRGLTGPFGPPGAGAPGDKGSGSPGAPGPTGARGAPGDRGEPGPPGAPGAPG 817  
 QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGPFGKAGRGVPPGPPGA---- 341  
 DB 818 GADGQPKAGEPDGAKGADAGPPGAPGAPGPIGNVGAFCAGKARGSPGATGPP 877  
 QY 342 -----VGPAGKDEAGAGQPPGAPG-----AGERGEGPAGSPGFGQLPPGAPGGEAGKP 393  
 DB 878 GAAGRVPFPSPGNAGPPGPPGAGKEGKGPRGETGAPRGCEVGPVPPGPPGAPGKSP 937  
 QY 394 GEQGVFDLCAAPSPGAPGPPGPTGLPPPPGRRGGSGRFPAGDGVAGKGPAGRGSP 453  
 DB 938 GADGAPAGPTFPQGIAGQGVVGLPQGRGRRFPGLPSPGEGPKQFSGASGERGPP 997  
 QY 454 GPAGP---KSGPGEAGRPGEAGLPAGKGLTSGSPGPDGKTTPPGAGDGRPPGPP 510

```
Db 998 GPMGPPCLAGPPGSEGEAGPAGGSGRDSFGAKXDRGETGPGAGPPGAPGAPGV 1057
Qy 511 GARGQAGVMGFPKGAAGBPGKAGRGVDPGPPGAVGPAGKQGE-----AGAQ 558
Db 1058 GPAGSGDRGETGAPGAPGVGAPGARGPAGPQGRDKGETGEGEQDRIKXHRGFSGLQ 1117
Qy 559 GPPGP-----AGPAGERGEOGPAGS-----GFQGLPGPAGPGEAGKPGGQGV 603
Db 1118 GPPGPPSGEOQPSGNSGAPGRGPPGSAGAPCKDGLNGLPFIPIGPPRGTGAGPV 1177
Qy 604 GDLGAPGSPGAP 616
Db 1178 GPPGPPGPPGPPG 1190

RESULT 6
US-10-357-851-1
; Sequence 1, Application US/10357851
; Publication No. US20040151731A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; TITLE OF INVENTION: Method and Compositions Involving
; TITLE OF INVENTION: T-Lymphocyte Reactivity with Collagen in Aortic Tissue of
; TITLE OF INVENTION: Abdominal Aortic Aneurysm Patients
; FILE REFERENCE: 13376US
; CURRENT APPLICATION NUMBER: US/10/357,851
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-357-851-1

Query Match 64.1%; Score 2237.5; DB 16; Length 1464;
Best Local Similarity 60.2%; Pred. No. 2.4e-104;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

Qy 1 GPP-----GPPGPTGLPCCPGERGSGRFPAGDGVAGPKGAGRGSPGPA 48
Db 458 GPPGAGEGKRGARGEPGPTGLPGERGGSPGRFPAGDGVAGPKGAGRGSPGPA 517
Qy 49 GPKGSPGAGRPAGAGLPGAKGLTSGSPGPDGKTGTPAGQDGRPGPPPGARGQA 108
Db 518 GPKGSPGAGRPAGAGLPGAKGLTSGSPGPDGKTGTPAGQDGRPGPPPGARGQA 577
Qy 109 GVMGPPGKGAAGBPGKAGRGVPPGAVGAPGAKDGEAGAGQPPGAPGAGERGEOGPA 168
Db 578 GVMGPPGKGAAGBPGKAGRGVPPGAVGAPGAKDGEAGAGQPPGAPGAGERGEOGPA 637
Qy 169 GSPGFQGLPGPAGPGEAGKPGEOQVPGDLGAPGSPGAGE-----PGP----- 212
Db 638 GSPGFQGLPGPAGPGEAGKPGEOQVPGDLGAPGSPGAGERGFPGERGVQPPGAPGR 697
Qy 213 -----TGLPGERGSGRFPAGDGVAGPKGAPGAGERGSPGAPKQSP 258
Db 698 GANGAPGNDGAKDAGAPGAGSQAPGLQMPGERGAAGLPGKGDGRDAGPKGADGSP 757
Qy 259 GE-----AGRPGEAGLPGAK-----GLTSPGSPGPDGKTGTPGAPQDGRP 300
Db 758 GKDGVRGLTGPIGPPGAPAGDCKGESGSPGAPGTGARGAPGRGREGPPGAPGAP 817
Qy 301 GPPGPPGARGQ-----AGVMGFPKGAAGBPGKAGRGVPPGPA----- 341
Db 818 GADGQFGAKGEPGDAGAKDAGPPGAPGAPGPPGPIGNVGAAPGAKGARGSPGATGTFP 877
Qy 342 -----VGPAGKDEAGAGPPGAPG-----AGERGEOGPAFGFQGLPGAPGPPGAGK 393
Db 878 GAAGRVGPPGSGNAGPPGPPGAPGKGGKGRGETGAGPGEVGGPPGPPGAGKQSP 937
Qy 394 GEQGVGDLGAPGSPGAGEPGPTGLPPOPGERGSGRFPAGDGVAGPKGAPGAGERGSP 453
```

```
Db 938 GADGPAGAPGTGPGQIAGQGVVGLPGQGERGFPGLPGSPGSEFGKQGPSASGERGPP 997
Qy 454 GPAGP---KGSPPGAGRGAGLPGAKGLTSGSPGPDGKTGTPGPPAGQDGRPGPPGP 510
Db 998 GPMGPPCLAGPPGSEGEAGPAGGSGRDSFGAKXDRGETGPGAGPPGAPGAPGV 1057
Qy 511 GARGQAGVMGFPKGAAGBPGKAGRGVDPGPPGAVGPAGKQGE-----AGAQ 558
Db 1058 GPAGSGDRGETGAPGAPGVGAPGARGPAGPQGRDKGETGEGEQDRIKXHRGFSGLQ 1117
Qy 559 GPPGP-----AGPAGERGEOGPAGS-----GFQGLPGPAGPGEAGKPGGQGV 603
Db 1118 GPPGPPSGEOQPSGNSGAPGRGPPGSAGAPCKDGLNGLPFIPIGPPRGTGAGPV 1177
Qy 604 GDLGAPGSPGAP 616
Db 1178 GPPGPPGPPGPPG 1190

RESULT 7
US-10-358-024-1
; Sequence 1, Application US/10358024
; Publication No. US20040151732A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; TITLE OF INVENTION: Method and Compositions Involving Blood
; TITLE OF INVENTION: T-Lymphocyte Reactivity with Collagen
; FILE REFERENCE: 13436US
; CURRENT APPLICATION NUMBER: US/10/358,024
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-358-024-1

Query Match 64.1%; Score 2237.5; DB 16; Length 1464;
Best Local Similarity 60.2%; Pred. No. 2.4e-104;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

Qy 1 GPP-----GPPGPTGLPCCPGERGSGRFPAGDGVAGPKGAGRGSPGPA 48
Db 458 GPPGAGEGKRGARGEPGPTGLPGERGGSPGRFPAGDGVAGPKGAGRGSPGPA 517
Qy 49 GPKGSPGAGRPAGAGLPGAKGLTSGSPGPDGKTGTPGAPQDGRPGPPPGARGQA 108
Db 518 GPKGSPGAGRPAGAGLPGAKGLTSGSPGPDGKTGTPGAPQDGRPGPPPGARGQA 577
Qy 109 GVMGPPGKGAAGBPGKAGRGVPPGAVGAPGAKDGEAGAGQPPGAPGAGERGEOGPA 168
Db 578 GVMGPPGKGAAGBPGKAGRGVPPGAVGAPGAKDGEAGAGQPPGAPGAGERGEOGPA 637
Qy 169 GSPGFQGLPGPAGPGEAGKPGEOQVPGDLGAPGSPGAGE-----PGP----- 212
Db 638 GSPGFQGLPGPAGPGEAGKPGEOQVPGDLGAPGSPGAGERGFPGERGVQPPGAPGR 697
Qy 213 -----TGLPGERGSGRFPAGDGVAGPKGAPGAGERGSPGAPKQSP 258
Db 698 GANGAPGNDGAKDAGAPGAGSQAPGLQMPGERGAAGLPGKGDGRDAGPKGADGSP 757
Qy 259 GE-----AGRPGEAGLPGAK-----GLTSPGSPGPDGKTGTPGAPQDGRP 300
Db 758 GKDGVRGLTGPIGPPGAPAGDCKGESGSPGAPGTGARGAPGRGREGPPGAPGAP 817
Qy 301 GPPGPPGARGQ-----AGVMGFPKGAAGBPGKAGRGVPPGPA----- 341
Db 818 GADGQFGAKGEPGDAGAKDAGPPGAPGAPGPPGPIGNVGAAPGAKGARGSPGATGTFP 877
Qy 342 -----VGPAGKDEAGAGPPGAPG-----AGERGEOGPAFGFQGLPGAPGPPGAGK 393
```

Db 878 GAAGRVGPPGSGNAGPPGPPAGKEGKGRGTGPACRGVGVGPPGPPAGEKGP 937  
 QY 394 GEQGVGDLGAPGSGPAGRGPTGLPDPGERGGPSRGFPAGADVAGVKGPAGERGSP 453  
 Db 938 GADGAPAGPTGPGQGIAGQGVVGLPQQRGERFPLPGPSGPGKQSGSGRGGPP 997  
 QY 454 GPAGP---KGSPPGACRGEAGLPGAKGLTSGPSGPDGKTGPPGAGDGRPPGPP 510  
 Db 998 GPMGPPGLAGPPGSGREGAGSGPDRGSPAKGDRGETGPAGPGAPGAPGVP 1057  
 QY 511 GARGQAVGVMGPPGKGAAGEFGKAGERGVPPGPAVGAGKDG------AGAQ 558  
 Db 1058 GPAGKSGDRGETGPAGPAGPVGAPGAGPAGPQGRGDKGTGEBQGDRIKGRHGFSLQ 1117  
 QY 559 GPPGP-----AGPAGERGEGPAGSP---GFQGLPGPAGPPGAGKPGGQVP 603  
 Db 1118 GPPGPPGSGQSGSGAGPAGPPGSGAGPKDGLNLGPIGPPGRGTGDAGPV 1177  
 QY 604 GDLGAPGSPGAG 616  
 Db 1178 GPPGPPGPPGPPG 1190

## RESULT 8

US-10-788-792-150  
 ; Sequence 150, Application US/10788792  
 ; Publication No. US20040191819A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bayer Pharmaceuticals Corporation  
 ; APPLICANT: Eveleigh, Douglas  
 ; APPLICANT: Bigwood, Douglas  
 ; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE  
 ; FILE REFERENCE: 5152  
 ; CURRENT APPLICATION NUMBER: US/10788,792  
 ; CURRENT FILING DATE: 2004-02-27  
 ; PRIOR APPLICATION NUMBER: US 60/450,655  
 ; PRIOR FILING DATE: 2003-02-28  
 ; NUMBER OF SEQ ID NOS: 254  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 150  
 ; LENGTH: 1464  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-788-792-150

Query Match 64.1%; Score 2237.5; DB 17; Length 1464;  
 Best Local Similarity 60.2%; Pred. No. 2.4e-104;  
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;  
 QY 1 GPP-----GPPGTGLPGRGGRGSGRFGPCADGVAGPKPAGERGSPGA 48  
 Db 458 GPPCPAGEEKGARGEPGTGLPGRGGRGSGRFGPCADGVAGPKPAGERGSPGA 517  
 QY 49 GPKSGPEAGKPEAGLPGAKGLTSGPSGPDGKTGPPGAGDGRPPGPPGARGQA 108  
 Db 518 GPKSGPEAGKPEAGLPGAKGLTSGPSGPDGKTGPPGAGDGRPPGPPGARGQA 577  
 QY 109 GVWGFPPKGAAGEPGKAGERGVGPPGAVGAPKDGGEAGAGQPPGAPGAGERGQGA 168  
 Db 578 GVWGFPPKGAAGEPGKAGERGVGPPGAVGAPKDGGEAGAGQPPGAPGAGERGQGA 637  
 QY 169 GSPGFQGLPGPAGPPGAGKPGGQGVFDLGAPEGSGPAGE-----PG- 212  
 Db 638 GSPGFQGLPGPAGPPGAGKPGGQGVFDLGAPEGSGPAGE-----PG- 212  
 QY 213 -----TGLPGRGGRGSGRFGPCADGVAGPKPAGERGSPGAPGKSP 258  
 Db 698 GANGAPNDGAKDAGAPGAPGQGAFLQGMFGERGAAGLPGPKDGRDAGPKGADSP 757  
 QY 259 GE-----AGRPGEAGLPGAK-----GLTSGPSGSPGDPGKTGPPGAGDGRP 300  
 Db 758 GKDGVRLGTGPIGPPGAPAGDKGSGSGPAGPTGARGAPCDRGEPPGPPAGFAGPP 817

QY 301 GPPGPPGARGQ-----AGVMGPPGPKGAAGEPKKAGERGVPPGPA----- 341  
 Db 818 GADGQFGAKGEBPDAGAKXDAGPPGAPGAPGPPGIGNVGAPKAGKAGSGAGPPGATGFP 877  
 QY 342 -----VGPAKDGEAGAGQPPGAPG---AGERBQGPAGSPGQGLPGPAGPPGAEAGKP 393  
 Db 878 GAAGRVGPPGSGNAGPPGPPGACKEGKGRGTGPAGRGVGVGPPGPPAGEKGP 937  
 QY 394 GEQGVGDLGAPGSGPAGRGPTGLPDPGERGGPSRGFPAGADVAGVKGPAGERGSP 453  
 Db 938 GADGAPAGPTGPGQGIAGQGVVGLPQQRGERFPLPGPSGPGKQSGSGRGGPP 997  
 QY 454 GPAGP---KGSPPGACRGEAGLPGAKGLTSGPSGPDGKTGPPGAGDGRPPGPP 510  
 Db 998 GPMGPPGLAGPPGSGREGAGSGPDRGSPAKGDRGETGPAGPGAPGAPGVP 1057  
 QY 511 GARGQAVGVMGPPGKGAAGEFGKAGERGVPPGPAVGAGKDG------AGAQ 558  
 Db 1058 GPAGKSGDRGETGPAGPAGPVGAPGAGPAGPQGRGDKGTGEBQGDRIKGRHGFSLQ 1117  
 QY 559 GPPGP-----AGPAGERGEGPAGSP---GFQGLPGPAGPPGAGKPGGQVP 603  
 Db 1118 GPPGPPGSGQSGSGAGPAGPPGSGAGPKDGLNLGPIGPPGRGTGDAGPV 1177  
 QY 604 GDLGAPGSPGAG 616  
 Db 1178 GPPGPPGPPGPPG 1190

## RESULT 9

US-10-402-089-8  
 ; Sequence 8, Application US/10402089  
 ; Publication No. US20040005663A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bell, Marcum P.  
 ; APPLICANT: Neff, Thomas B.  
 ; APPLICANT: Polarek, James W.  
 ; APPLICANT: Seeley, Todd W.  
 ; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS  
 ; FILE REFERENCE: FP0402.3 CON  
 ; CURRENT APPLICATION NUMBER: US/10/402,089  
 ; CURRENT FILING DATE: 2003-03-26  
 ; PRIOR APPLICATION NUMBER: US 09/709,700  
 ; PRIOR FILING DATE: 2000-11-10  
 ; NUMBER OF SEQ ID NOS: 72  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 8  
 ; LENGTH: 1449  
 ; TYPE: PRT  
 ; ORGANISM: Sus scrofa  
 US-10-402-089-8

Query Match 64.0%; Score 2234; DB 15; Length 1449;  
 Best Local Similarity 62.0%; Pred. No. 3.6e-104;  
 Matches 438; Conservative 21; Mismatches 154; Indels 94; Gaps 13;  
 QY 1 GPPGEPGTGLPGRGGRGSGRFGPCADGVAGPKPAGERGSPGAPGKSGPGEAGRP 60  
 Db 472 GARGEPGAPGLPGPPGGRGSGRFGPCADGVAGPKPAGERGSPGAPGKSGPGEAGRP 531  
 QY 61 GEAGLPGAKGLTSGPSGPDGKTGPPGAGDGRPPGPPGARGQAVGVMGPPGKGA 120  
 Db 532 GEAGLPGAKGLTSGPSGPDGKTGPPGAGDGRPPGPPGARGQAVGVMGPPGKGA 591  
 QY 121 GEPKAGERGVPPGPAVGAPKDGGEAGAGQPPGAPGAGERGQGPAGSPGQGLPGPA 180  
 Db 592 GEPKAGERGVPPGPAVGAPKDGGEAGAGQPPGAPGAGERGQGPAGSPGQGLPGPA 651  
 QY 181 GPPGAEAKPGEQGVFDLGAPEGSGPAGE-----PG- 212  
 Db 652 GPPGAEAKPGEQGVFDLGAPEGSGPAGE-----PG- 212  
 QY 213 --TCLPDPGERGSGRFGPCADGVAGPKPAGERGSPGAPGKSGPGE-----A 261

Db 712 GDAGAPAPGSGAGLQGVGGERGAGLPGPKDRGDAGPKADGAPKDGVRGLTPI 771  
Qy 262 GRPGEAGLPGAKGLTSGSGPBDGKTGPPGAGQDGRGP-PPGPGARGAQVWGF- 317  
Db 772 GPPGAPAGDGTGSGGAGTARGAPQDRGEPGPPGAGFAGPPGADGQFAGKGGP 831  
Qy 318 ---PGPKGAGEPKKAGRGVGP-PPGA-----VGPAGKDGAGAGQPPGAPAGE 365  
Db 832 TGPPGTGSGVAGPAPKARGAGPPGATGPPGAGRVGPPGPGSNAGPPGPPGAGKEGS 891  
Qy 366 ---RGEQGPAGSPGQGLPGPAGPPGAGKPGEGQGVGDLGAPGSPGAGEPPTGLPGP 422  
Db 892 KGRGEGTGPAGRPGGAGPPGPPGAGEKSGFADGAPAGTTPGQGIAGQGVVGLPGQ 951  
Qy 423 PBERGPGSRGFPAGDVGAGPKGAGRGSGPPGAPKSGPCEAGRPGKGLTGS 482  
Db 952 RGERGFPGLPGSPGEPKQKSGSGSGRGP-PPGMP-PCLAGPPSGREGAPGAEGS 1008  
Qy 483 P---GSPGPD---GKTGPPGAGQDGRPPGPPGARGQAGVGFPPKGAAGEPPKAGE 536  
Db 1009 PGRDGAFCPKGDRGESGAPGAPGAPGAPGVPVGPAGKSGDRGCTGTPAGPAGVPVGA 1068  
Qy 537 RGVPPGAPGAVPAGKDE-----AGAGPPGP-----AGPAGERGE 572  
Db 1069 RGPAGPQGRGDKGTGEGQDRGKIKHGRGSLGQPPGPPGSGPGEQPSGASGAPGRGP 1128  
Qy 573 QGPAGSP---CFQGLPGPAGPPGAGKPGEGQGVGDLGAPGSPGAPG 616  
Db 1129 PGSAGAPGKGLNGLPPIGPPGRTGAGPVGPPGPPGPPG 1175

RESULT 10  
US-10-402-072A-8  
; Sequence 8, Application US/10402072A  
; Publication No. US20040018592A1  
; GENERAL INFORMATION:  
; APPLICANT: Bell, Marcum P.  
; APPLICANT: Neff, Thomas B.  
; APPLICANT: Polarek, James W.  
; APPLICANT: Seeley, Todd W.  
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS  
; FILE REFERENCE: EP0402.2 CON  
; CURRENT APPLICATION NUMBER: US/10/402,072A  
; CURRENT FILING DATE: 2003-03-26  
; PRIOR APPLICATION NUMBER: US 09/709,700  
; PRIOR FILING DATE: 2000-11-10  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 8  
; LENGTH: 1449  
; TYPE: PRT  
; ORGANISM: Sus scrofa  
US-10-402-072A-8

Query Match 64.0%; Score 2234; DB 15; Length 1449;  
Best Local Similarity 52.0%; Pred. No. 3.6e-104; Indels 94; Gaps 13;  
Matches 438; Conservative 21; Mismatches 154;  
Qy 1 GPPGPPGTGLPFPGERGPPGAGDVGAGPKGAGRGSGPPGAGKSGPCEAGR 60  
Db 472 GARGEPGAGLPGPPGERGPPGAGDVGAGPKGAGRGSGPPGAGKSGPCEAGR 531  
Qy 61 GEAGLPGAKGLTSGSGPDPGKTGPPGAGDGRGPPGPPGARGQAGVGFPGKGA 120  
Db 532 GEAGLPGAKGLTSGSGPDPGKTGPPGAGDGRGPPGPPGARGQAGVGFPGKGA 591  
Qy 121 GPPGKAGRGVPPGPGVAGPKGAGEAGAQPPGAPGAGRGQGPAGSGPGLPGPA 180  
Db 592 GEPGKAGRGVPPGPGVAGPKGAGEAGAQPPGAPGAGRGQGPAGSGPGLPGPA 651  
Qy 181 GPPGKAGRGVPPGDLGAPGSPGAGE-----PGP----- 212

Db 652 GPPGAGKPGEGQGVGDLGAPGSPGARGGPPGPPGQGVGPPGAGPPGANGAPGNDGAK 711  
Qy 213 ---TGLPGPPBERGSGSGFQADVAGPKGAPGAGRGSGPPGAGKSGPGE-----A 261  
Db 712 GDAGAPAPGSGAGLQGVGGERGAGLPGPKDRGDAGPKADGAPKDGVRGLTPI 771  
Qy 262 GRPGEAGLPGAKGLTSGSGPBDGKTGPPGAGQDGRGP-PPGPGARGAQVWGF- 317  
Db 772 GPPGAPAGDGTGSGGAGTARGAPQDRGEPGPPGAGFAGPPGADGQFAGKGGP 831  
Qy 318 ---PGPKGAGEPKKAGRGVGP-PPGA-----VGPAGKDGAGAGQPPGAPAGE 365  
Db 832 TGPPGTGSGVAGPAPKARGAGPPGATGPPGAGRVGPPGPGSNAGPPGPPGAGKEGS 891  
Qy 366 ---RGEQGPAGSPGQGLPGPAGPPGAGKPGEGQGVGDLGAPGSPGAGEPPTGLPGP 422  
Db 892 KGRGEGTGPAGRPGGAGPPGPPGAGEKSGFADGAPAGTTPGQGIAGQGVVGLPGQ 951  
Qy 423 PBERGPGSRGFPAGDVGAGPKGAGRGSGPPGAPKSGPCEAGRPGKGLTGS 482  
Db 952 RGERGFPGLPGSPGEPKQKSGSGSGRGP-PPGMP-PCLAGPPSGREGAPGAEGS 1008  
Qy 483 P---GSPGPD---GKTGPPGAGQDGRPPGPPGARGQAGVGFPPKGAAGEPPKAGE 536  
Db 1009 PGRDGAFCPKGDRGESGAPGAPGAPGAPGVPVGPAGKSGDRGCTGTPAGPAGVPVGA 1068  
Qy 537 RGVPPGAPGAVPAGKDE-----AGAGPPGP-----AGPAGERGE 572  
Db 1069 RGPAGPQGRGDKGTGEGQDRGKIKHGRGSLGQPPGPPGSGPGEQPSGASGAPGRGP 1128  
Qy 573 QGPAGSP---GFQGLPGPAGPPGAGKPGEGQGVGDLGAPGSPGAPG 616  
Db 1129 PGSAGAPGKGLNGLPPIGPPGRTGAGPVGPPGPPGPPG 1175

RESULT 11  
US-10-468-091-25  
; Sequence 25, Application US/10468091  
; Publication No. US20040157329A1  
; GENERAL INFORMATION:  
; APPLICANT: ADP Pharmaceutical Pty Limited  
; APPLICANT: The University of Sydney  
; TITLE OF INVENTION: Matrix gene expression in chondrogenesis  
; FILE REFERENCE: 500311  
; CURRENT APPLICATION NUMBER: US/10/468,091  
; CURRENT FILING DATE: 2003-08-13  
; PRIOR APPLICATION NUMBER: AU PR3116  
; PRIOR FILING DATE: 2001-02-15  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 25  
; LENGTH: 1461  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-468-091-25

Query Match 64.0%; Score 2233.5; DB 16; Length 1461;  
Best Local Similarity 60.0%; Pred. No. 3.8e-104; Indels 117; Gaps 12;  
Matches 440; Conservative 23; Mismatches 153;  
Qy 1 GPP-----GEPGPTGLPFPGERGPPGAGDVGAGPKGAPGAGRGSGPPGA 48  
Db 455 GPPGAGEGKRGARGEPGPTGLPFPGERGPPGAGDVGAGPKGAPGAGRGSGPPGA 514  
Qy 49 GPKGSGEAGRPGKAGLPGAKGLTSGSGPDPGKTGPPGAGDGRGPPGPPGARGQA 108  
Db 515 GPKGSGEAGRPGKAGLPGAKGLTSGSGPDPGKTGPPGAGDGRGPPGPPGARGQA 574  
Qy 109 GVMGFPKPGKAGEPKKAGRGVPPGPPGAVGPPAGKDGAGEAGAQPPGAPGAGERGQGA 168  
Db 575 GVMGFPKPGKAGEPKKAGRGVPPGPPGAVGPPAGKDGAGEAGAQPPGAPGAGERGQGA 634  
Qy 169 GSPGFGQLPGPAGPPGKAGEQGVGDLGAPGSPGAGE-----PGP----- 212

[illegible]

```

RESULT 12
US-10-291-265-243
; Sequence 243, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291.265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: Fast-SEQ for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-243

```

[illegible]

RESULT 13  
US-10-104-889-16  
Sequence 16, Application US/10104889  
Publication No. US20040086961A1  
GENERAL INFORMATION:  
APPLICANT: GRUSKIN, ELLIOT A.  
BUECHTER, DOUGLAS  
BROKAW, JANE  
ZHANG, GUANGHUI  
PAOLELLA, DAVID  
TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DILWORTH & BARRESE  
STREET: 333 EARLE OVINGTON BOULEVARD  
CITY: UNIONDALE  
STATE: NY  
COUNTRY: U.S.A.  
ZIP: 11553  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104.889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-10-104-889-16

Query Match      64.0%; Score 2231.5; DB 15; Length 1057;
Best Local Similarity 60.0%; Pred. No. 3.8e-104;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GPGPTGLPGPPGGRGGPGSGRFPAGDVGAGPKGAGRGSGPGPA 48
DB 297 GPPGAGEEGRGARGEGFTGLPGPPGGRGGPGSGRFPAGDVGAGPKGAGRGSGPGPA 356
QY 49 GPKSGPEAGRPBAGLPGAKGLTSGPSGPGDGTGPPAGDQGRPPGPPGARGQA 108
DB 357 GPKSGPEAGRPBAGLPGAKGLTSGPSGPGDGTGPPAGDQGRPPGPPGARGQA 416
QY 109 GWMGFPQKGAAGBPGKAGRGVPPGAVGPAKDGAGAAQPPGAGPAGRGSGGPA 168
DB 417 GWMGFPQKGAAGBPGKAGRGVPPGAVGPAKDGAGAAQPPGAGPAGRGSGGPA 476
QY 169 GSRGFGQLPGPAGPBGAGKEGEGVGDLCAGPSPGAGE-----PGP---- 212
DB 477 GSPGFGQLPGPAGPBGAGKEGEGVGDLCAGPSPGARGERFPGERGVQGPAGPR 536
QY 213 -----TGLPFPGERGGPGSGRFPAGDVGAGPKGAGRGSGPGPKGSP 258
DB 537 GANGAPGNDGAKGDAGAPGAGSGQAPGLQMPGECGAGLPGPKGDRGDAGPKGADGSP 596
QY 259 GE-----AGRPGENGLPKAK-----CLTSGPSGPGDGTGPPGAGQDGRP 300
DB 597 GKDGVRLGTGTPGPPGAPAGFGDKGESGPGAGFTGARGAPGDRGEPGPPGAGFAGPP 656
QY 301 GPPGPPGARGQ-----AGVMGFPQKGAAGBPGKAGRGVFPGPGA----- 341
DB 657 GADGQPGAKGFPGDAGAKGDAGFPAGPAGPGPIGNVGAPGAKGARGAGPPGATGFP 716
QY 342 -----VGPAGKDGAGAGGPPGAPG---AGERGQGPAGSPGQGLPGPAGPPGEGAKP 393
DB 717 CAACRVGPPGPGSGNAGPPGPGKAGCKGKGRGTGAPRGFVGPPGPPGAGRGSP 776
QY 394 GEQGVGDLGAPGSGPAGPGTGLPGPPGGRGGPGSRFPAGDVGAGPKGAGRGSP 453
DB 777 GADGAPAGPTGPGFGIAGQGVVLPGQRGERFPLPGPSGPPGQKQSGSAGSGERKPP 836
QY 454 GPAGS---KGSFGEAGRGEAGLPGAKGLTSGPSGPGDGTGPPGAGQDGRPPGPPG 510
DB 837 GPMGPPGLAGFPGESREGAPAAESGSPGRDGSPPGAKGDRGTGTPAGPPGAPGAPGVP 896
QY 511 GARGQAGVMGFPFGPKGAAGECKKAGRCVPPGPAVGPAGKDG-----ACAG 558
DB 897 GPAGKSGDRGTGTGAPGAPGVPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 956
QY 559 GPPGP-----AGPAGERGEQGPAGSP---GFGQLPGPAGPPGAGKPGEGQVP 603

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104.889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-10-104-889-16

Query Match      63.9%; Score 2228.5; DB 14; Length 1341;
Best Local Similarity 60.0%; Pred. No. 6.4e-104;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GPGPTGLPGPPGGRGGPGSGRFPAGDVGAGPKGAGRGSGPGPA 48
DB 334 GPPGAGEEGRGARGEGFTGLPGPPGGRGGPGSGRFPAGDVGAGPKGAGRGSGPGPA 393
QY 49 GPKSGPEAGRPBAGLPGAKGLTSGPSGPGDGTGPPGAGQDGRPPGPPGARGQA 108
DB 394 GPKSGPEAGRPBAGLPGAKGLTSGPSGPGDGTGPPGAGQDGRPPGPPGARGQA 453

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/058,124
; FILING DATE: 29-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/570,573
; FILING DATE: 2002-MAY-12
; APPLICATION NUMBER: 08/187,319
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: GOSORIS, ADDA C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-058-124-18

Query Match      63.9%; Score 2228.5; DB 14; Length 1341;
Best Local Similarity 60.0%; Pred. No. 6.4e-104;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GPGPTGLPGPPGGRGGPGSGRFPAGDVGAGPKGAGRGSGPGPA 48
DB 334 GPPGAGEEGRGARGEGFTGLPGPPGGRGGPGSGRFPAGDVGAGPKGAGRGSGPGPA 393
QY 49 GPKSGPEAGRPBAGLPGAKGLTSGPSGPGDGTGPPGAGQDGRPPGPPGARGQA 108
DB 394 GPKSGPEAGRPBAGLPGAKGLTSGPSGPGDGTGPPGAGQDGRPPGPPGARGQA 453
```

QY 109 GVMGPPGKGAAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 168  
 DB 454 GVMGPPGKGAAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 513  
 QY 169 GSPGQGLPAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 212  
 DB 514 GSPGQGLPAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 573  
 QY 213 -----TGLPGERGPGSRGPPGADGAGAGAAQPPGPPGAGPAGERGEOGPA 258  
 DB 574 GANCAPNDGAKGAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 633  
 QY 259 GE-----AGPGEAGLPGAK-----GLTSGPSGPPGDKTGTGPPGAGPAGERGEOGPA 300  
 DB 634 CKDQVRLGTGPIGPPGAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 693  
 QY 301 GPPGPPGARGQ-----AGVMPGPKGAAGPCKAGRGVPGPPGAGPAGERGEOGPA 341  
 DB 694 GADQAPGAKGAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 753  
 QY 342 -----VGPAGKDGAGAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 393  
 DB 754 GAAGRVGPPGSGNAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 813  
 QY 394 GEQGVGDLGAPGSGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 453  
 DB 814 GADQAPGAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 873  
 QY 454 GPAGP-----KSGPAGKDGAGAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 510  
 DB 874 GPMGPPGLAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 933  
 QY 511 GARGQAGVMPGPPGKGAAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 558  
 DB 934 GPAGKSGDRGTGTGAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 993  
 QY 559 GPPGPP-----AGPAGERGEOGPA 603  
 DB 994 GPPGPPSGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 1053  
 QY 604 GDLGAPGSPGAG 616  
 DB 1054 GPPGPPGPPGPPG 1066

## RESULT 15

US-09-918-715-261  
 ; Sequence 261, Application US/09918715  
 ; Publication No. US20030017157A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brad St. Croix  
 ; APPLICANT: Bert Vogelstein  
 ; APPLICANT: Kenneth Kinzler  
 ; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
 ; FILE REFERENCE: 1107.00134  
 ; CURRENT APPLICATION NUMBER: US/09/918,715  
 ; CURRENT FILING DATE: 2001-08-01  
 ; PRIOR APPLICATION NUMBER: 60/222,599  
 ; PRIOR FILING DATE: 2000-08-02  
 ; PRIOR APPLICATION NUMBER: 60/224,360  
 ; PRIOR FILING DATE: 2000-08-11  
 ; PRIOR APPLICATION NUMBER: 60/282,850  
 ; PRIOR FILING DATE: 2000-04-11  
 ; NUMBER OF SEQ ID NOS: 358  
 ; SOFTWARE: Fast-Seq for Windows Version 3.0  
 ; SEQ ID NO 261  
 ; LENGTH: 1464  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-918-715-261

Query Match

63.9%; Score 2227.5; DB 10; Length 1464;

Best Local Similarity 59.9%; Pred. No. 7.7e-104;  
 Matches 439; Conservative 23; Mismatches 154; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGERGPGSRGPPGADGAGAGAAQPPGPPGAGPAGERGEOGPA 48  
 DB 458 GPPGAGBEGKRGKAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 517  
 QY 49 GPKGSPGAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 108  
 DB 518 GPKGSPGAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 577  
 QY 109 GVMGPPGKGAAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 168  
 DB 578 GVMGPPGKGAAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 637  
 QY 169 GSPGQGLPAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 212  
 DB 638 GSPGQGLPAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 697  
 QY 213 -----TGLPGERGPGSRGPPGADGAGAGAAQPPGPPGAGPAGERGEOGPA 258  
 DB 698 GANCAPNDGAKGAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 757  
 QY 259 GE-----AGPGEAGLPGAK-----GLTSGPSGPPGDKTGTGPPGAGPAGERGEOGPA 300  
 DB 758 CKDQVRLGTGPIGPPGAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 817  
 QY 301 GPPGPPGARGQ-----AGVMPGPPGKGAAGPCKAGRGVPGPPGAGPAGERGEOGPA 341  
 DB 818 GADQAPGAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 877  
 QY 342 -----VGPAGKDGAGAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 393  
 DB 878 GAAGRVGPPGSGNAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 937  
 QY 394 GEQGVGDLGAPGSGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 453  
 DB 938 GADQAPGAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 997  
 QY 454 GPAGP-----KSGPAGKDGAGAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 510  
 DB 998 GPMGPPGLAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 1057  
 QY 511 GARGQAGVMPGPPGKGAAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 558  
 DB 1058 GPAGKSGDRGTGTGAGPAGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 1117  
 QY 559 GPPGPP-----AGPAGERGEOGPA 603  
 DB 1118 GPPGPPSGPCKAGRGVPGPPGAVGPAKDGAGAGAAQPPGPPGAGPAGERGEOGPA 1177  
 QY 604 GDLGAPGSPGAG 616  
 DB 1178 GPPGPPGPPGPPG 1190

Search completed: October 18, 2004, 13:58:28  
 Job time : 84.3776 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:31:10 ; Search time 24.4569 Seconds  
(without alignments)  
2427.363 Million cell updates/sec

Title: US-10-658-989A-2

Perfect score: 3488  
Sequence: 1 GPPGEGPTGLPFGPGRGG.....GPGVFGDGLGAPGSPAGG 617

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:\*

1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2228.5	63.9	1464	1 CGHUIS	collagen alpha 1(I)
2	2205.5	63.2	1453	2 S21626	collagen alpha 1(I)
3	2151.5	61.7	1042	1 CGCHIS	collagen alpha 1(I)
4	2011.5	57.7	671	1 CGRTIS	collagen alpha 1(I)
5	1944	55.7	1418	2 T45467	collagen alpha 1(I)
6	1943	55.7	1487	1 CGHUCG	collagen alpha 1(I)
7	1932	55.4	1419	2 A41182	collagen alpha 1(I)
8	1932	55.4	1487	2 B41182	collagen alpha 1(I)
9	1898	54.4	779	1 CGBOIS	collagen alpha 1(I)
10	1890.5	54.2	1486	1 B40333	collagen alpha 1(I)
11	1883	54.0	1492	2 A40333	collagen alpha 1(I)
12	1857.5	53.3	1464	2 S58556	collagen alpha 1(I)
13	1856	53.2	1049	1 CGBO7S	collagen alpha 1(I)
14	1853	53.1	1466	1 CGHUVL	collagen alpha 1(I)
15	1840	52.8	886	2 I50694	collagen alpha 1(I)
16	1822.5	52.3	1496	1 CGHUV7	collagen alpha 2(V)
17	1804	51.7	1497	2 I49607	procollagen type V
18	1788.5	51.3	1373	1 A43291	collagen alpha 2(I)
19	1764	50.6	1365	1 CGHUVS	collagen alpha 2(I)
20	1702.5	48.8	1838	1 CGHUVI	collagen alpha 1(V)
21	1689.5	48.4	1843	2 S18803	collagen alpha 1(V)
22	1644	47.1	1806	1 CGHUIE	collagen alpha 1(X)
23	1640.5	47.0	1414	1 S23809	collagen alpha 2(I)
24	1612	46.2	1027	2 S28774	collagen alpha 2(I)
25	1587	45.5	1024	2 S18251	collagen alpha 1(X)
26	1585	45.4	1546	1 CGHUIE	collagen alpha 2(X)
27	1584	45.4	1691	1 S22917	collagen alpha 5(I)
28	1583.5	45.4	673	1 CGBO6C	collagen alpha 1(I)
29	1568	45.0	2944	2 A54849	collagen alpha 1(V)

30 1556.5 44.6 1549 2 148103 type VII collagen  
31 1550.5 44.5 888 2 S28791 collagen alpha 1(X)  
32 1547 44.4 1690 1 CGHUIB collagen alpha 4(I)  
33 1530 43.9 1763 2 S16366 collagen alpha 2(I)  
34 1526 43.8 964 1 CGCH2S collagen alpha 2(I)  
35 1522 43.6 1669 1 CGHUA8 collagen alpha 1(I)  
36 1510 43.3 920 2 A45748 collagen alpha 1(V)  
37 1508 43.2 1603 2 S23810 collagen alpha 1(X)  
38 1503.5 43.1 1669 1 CGMS4B collagen alpha 1(I)  
39 1489.5 42.7 1670 1 CGHUIB collagen alpha 3(I)  
40 1481 42.5 1712 1 CGHUIB collagen alpha 2(I)  
41 1471.5 42.2 812 2 S31521 collagen COLF1 - f  
42 1463 41.9 1759 2 T29351 collagen alpha 2(I)  
43 1459.5 41.8 1758 2 T29350 hypothetical prote  
44 1456 41.7 1142 2 JX0369 collagen alpha 1(X)  
45 1437.5 41.2 1752 2 A45407 collagen alpha 3(I)

#### ALIGNMENTS

##### RESULT 1

CGHUIS

N/Alternate names: procollagen alpha 1(I) chain  
collagen alpha 1(I) chain precursor - human

C/Species: Homo sapiens (man)

C/Date: 12-Aug-1981 #sequence revision 04-Oct-1996 #text change 09-Jul-2004

C/Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S1:

5269; A29439; I53466; A02852; I37247

R/D/Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, P.; Pretorius, P.J.  
Gene 67, 105-115, 1988

A/Title: Complete nucleotide sequence of the region encompassing the first twenty-five  
A/Reference number: I60114; MUID:88329734; PMID:2843432

A/Accession: I60114

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-369, 'L', 371-589 <DAL>

A/Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNI  
R/Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Proc

Biochem. J. 253, 919-922, 1988

A/Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human  
A/Reference number: S01143; MUID:89025644; PMID:3178743

A/Accession: S01143

A/Molecule type: mRNA

A/Residues: 1-472 <TRO>

A/Cross-references: EMBL:X07884; NID:G30015; PIDN:CAA30731.1; PID:G30016; GB:M36546; NI  
A/Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988

R/Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.  
Nature 310, 337-340, 1984

A/Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation o  
A/Reference number: A93335; MUID:84270697; PMID:6462220

A/Accession: A93335

A/Molecule type: DNA

A/Residues: 1-58, 'Q', 60-181 <CHU>

A/Cross-references: EMBL:X00820; NID:G35657; PIDN:CAA25394.1; PID:G35658

R/Rossouw, C.M.S.; Vergeer, W.P.; du Picoy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.  
J. Biol. Chem. 262, 15151-15157, 1987

A/Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene en  
A/Reference number: I55254; MUID:88033098; PMID:2822714

A/Accession: I55254

A/Status: translation not shown; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-45 <ROS>

A/Cross-references: GB:J02829; NID:G180387; PIDN:AAA51993.1; PID:G180388

R/Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gellinas, R.E.

Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987

A/Title: Regulatory elements in the first intron contribute to transcriptional control  
A/Reference number: A39943; MUID:88097389; PMID:3480516

A/Accession: A39943

A/Molecule type: DNA

A/Residues: 1-34 <BOR>

A/Cross-references: GB:J03559; NID:G180876; PIDN:AAA52052.1; PID:G553238

R/Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

- J. Biol. Chem. 260, 2315-2320, 1985  
A>Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s  
A/Reference number: I55237; MUID:85130970; PMID:2857713  
A/Accession: I55237  
A>Status: translation not shown; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-34 <CH2>  
A/Cross-references: GB:M10627; NID:g180383; PIDN:AAA5192.1; PID:g53226  
R/Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist  
J. Biol. Chem. 265, 6312-6317, 1990  
A>Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina  
rome, type VII.  
A/Reference number: A35233; MUID:90202906; PMID:2318855  
A/Accession: A35233  
A/Molecule type: protein  
A/Residues: 33-52 <WR>  
A/Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved  
R/Well, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.  
EMBO J. 8, 1705-1710, 1989  
A>Title: A base substitution in the exon of a collagen gene causes alternative splicing  
A/Reference number: S09400; MUID:89356643; PMID:2767050  
A/Accession: S09400  
A/Molecule type: mRNA  
A/Residues: 156-183 <WEI>  
R/Click, E.M.; Bornstein, P.  
Biochemistry 9, 4699-4706, 1970  
A>Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1  
A/Reference number: A90567; MUID:71038625; PMID:529514  
A/Contents: CNBR0-1, CNBR2, CNBR4, CNBR5  
A/Accession: B90567  
A/Molecule type: protein  
A/Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'  
A/Experimental source: skin  
A/Note: evidence for 170-allysine  
R/Baetge, B.; Norbom, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutmann, R.; Mueller, F.  
Eur. J. Biochem. 192, 153-159, 1990  
A>Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle  
A/Reference number: S11372; MUID:90382436; PMID:2169412  
A/Accession: S11372  
A/Molecule type: protein  
A/Residues: 175-187, 274-287, 'P', 289 <BAE>  
A/Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion  
R/Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez  
J. Biol. Chem. 266, 21827-21832, 1991  
A>Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain  
operative melting of intact type I collagen.  
A/Reference number: I55342; MUID:92042092; PMID:1718984  
A/Accession: I55342  
A>Status: translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 258-268, 1347-1357 <DEA>  
A/Cross-references: GB:S67495; NID:g239007; PIDN:AA20350.1; PID:g239008  
A/Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report  
R/Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.  
J. Biol. Chem. 245, 5042-5048, 1970  
A>Title: Comparative study of glycosylated collagen derived from selected vertebrate collagens.  
A/Reference number: A92069; MUID:71001508; PMID:4319110  
A/Accession: A92069  
A/Molecule type: protein  
A/Residues: 263-268 <MOR>  
A/Experimental source: skin  
R/Labhard, M.E.; Hollister, D.W.  
Matrix 10, 124-130, 1990  
A>Title: Segmental amplification of the entire helical and telopeptide regions of the co  
A/Reference number: S15989; MUID:90326017; PMID:2374517  
A/Accession: S15989  
A/Molecule type: mRNA  
A/Residues: 281-302, 402-420, 823-843, 925-944, 1026-1045, 1143-1162 <LAB>  
R/Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N.  
Connect. Tissue Res. 29, 1-11, 1993  
A>Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of  
A/Reference number: I52905; MUID:93339042; PMID:8339541  
A/Accession: I52905

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1187-1194, 'C', 1196-1220 <COH>

A:Cross-references: GB:M23213; NID:G340842; PIDN:AAB59363.1; PID:G499622

A>Note: mutant sequence from a patient with mild osteogenesis imperfecta

R:Maekelae, J.K.; Raassina, M.; Virta, A.; Vuorio, E.

Nucleic Acids Res. 15, 349, 1988

A>Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

Query Match 63.9%; Score 2228.5; DB 1; Length 1464;

Best Local Similarity 60.0%; Pred. No. 4.2e-105;

Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

Qy 1 GPP-----GPGGTGLPGRGSGRFGPCADGVAGPKPAGRGSPGPA 48

Db 458 GPPGAGEGKRGARGEPGTGLPGRGSGRFGPCADGVAGPKPAGRGSPGPA 517

Qy 49 GPKGSPGAEAGPAGAGLPGAKGLTGTSGSPGPPGKGTGPPGAGQDGRPGPPPGARGQA 108

Db 518 GPKGSPGAEAGPAGAGLPGAKGLTGTSGSPGPPGKGTGPPGAGQDGRPGPPPGARGQA 577

Qy 109 GVMGFPDPKGAAGPAGAGRGVPPGPGAVGPAGKDGEGAGAQPPGAPGAGRGSGGPA 168

Db 578 GVMGFPDPKGAAGPAGAGRGVPPGPGAVGPAGKDGEGAGAQPPGAPGAGRGSGGPA 637

Qy 169 GSPGFGGLPAGPPGAGKPGEGVPGDLGAPGSPGAGE-----PGP----- 212

Db 638 GSPGFGGLPAGPPGAGKPGEGVPGDLGAPGSPGAGE-----PGP----- 697

Qy 213 -----TGLPGPGRGSGRFGPGADGVAGPKPAGRGSPGAGPKGSP 258

Db 698 GANGAPGNDGAKGADAGAPGSPGAGLQMPGEGRAAGLPKPGDRGAGPKGADGSP 757

Qy 259 GE-----AGPGEAGLPGAK-----CLTSGSPGSPDGKTPGPPGAGQDGRP 300

Db 758 GKGVRLGTGIPGPPGAGAPGDKGSGSPGSPGAGTGAAGAPGDRGPPGPPGAGPAGPP 817

Qy 301 GPPPPPPARGQ-----AGVMGFPFGKGAAGPAGKAGRGVPGPPGA----- 341

Db 818 GADQPGAKGPPGDAGAKGDAGPPGAPGAPGPIGNVGPAGKAGSAGPPGATGPP 877

Qy 342 -----VPGADKGBAGAGPPGAPG---AGERGQSPAGSPGQGLPAGPPGAGKXP 393

Db 878 GAAGRVPPGSPGNAGPPGPPGAGKGGKPGRGTGAPRGSEVGPSPGPPGAGKXP 937

Qy 394 GEQGVGDLGAPGSPGAGEPPTGLPGPGRGSGRFPGADGVAGPKPAGRGSP 453

Db 938 CADGPAGAPGTGPGTGTAGQGVVLPGQGRGTFPGLPGSPGPKQSGASGERGPP 997

Qy 454 GPAGP---XGSPGAGRGEAGLPGAKGLTSGSPGPDGKTPGPPGAGQDGRPPGPP 510

Db 998 GPMGPPGLAGPPGSGREGAPGAGSGPGRDGSAGKGRGTGTPAGPPGAXGAGPVP 1057

Qy 511 GARGQAGVMGPPGKGAAGEPKAGRGVPPGPGAVGPAGKDG-----AGAQ 558

Db 1058 GPAGSGDRGTGTPAGPAGVVPAGAPAGPQPRGDKGTGQGRGKIGHRGFSGLQ 1117

Qy 559 GPPGP-----AGPAGERGEQGPAGSP---QFQGLPAGPPGAGKPGEGQVP 603

Db 1118 GPPGPPSGGQSGASGAPGPPGPGSAGAPKQGLNGLPGIPGPPGRGTGDAGVP 1177

Qy 604 GDLGAPGSPGAG 616

Db 1178 GPPGPPGPPGPPG 1190

RESULT 2

S21626

collagen alpha 1(I) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text\_change 09-Jul-2004

C:Accession: S57243; S16374; A23982; I49559; S39789; I48300; S21626

R:Li, S.W.; Khillan, J.; Prockop, D.J.

Matrix Biol. 14, 593-595, 1994

A>Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I

A:Reference number: S57243

A:Accession: S57243

A:Molecule type: mRNA

A:Residues: 1-1453 <LIS>

A:Cross-references: UNIPROT:P11087; EMBL:U08020; NID:G470673; PIDN:AAA88912.1; PID:G470

R:Vetaserana, M.; Tonan, D.; de Crombrughe, B.; Vuorio, E.

Biochim. Biophys. Acta 1089, 241-243, 1991

A>Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.

A:Reference number: S16176; MUID:91274355; PMID:2054384

A:Accession: S16374

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1442-1453 <MET>

A:Cross-references: EMBL:X57981; NID:G50484; PIDN:CAA41046.1; PID:G50485

R:French, B.T.; Lee, W.H.; Maul, G.G.

Gene 39, 311-312, 1985

A>Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.

A:Reference number: A23982; MUID:86137403; PMID:3841523

A:Accession: A23982

A:Molecule type: mRNA

A:Residues: 518-1128 <PRE>

A:Cross-references: GB:M14423; NID:G192263; PIDN:AAA37333.1; PID:G192262

R:Monson, J.M.; Friedman, J.; McCarthy, B.J.

Mol. Cell. Biol. 2, 1362-1371, 1982

A>Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for

A:Reference number: I49559; MUID:83141374; PMID:6298597

A:Accession: I49559

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 735-1130 <RES>

A:Cross-references: GB:M17491; NID:G192263; PIDN:AAA37334.1; PID:G192264

R:Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.

Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984

A>Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads

A:Reference number: I49557; MUID:84170331; PMID:6324198

A:Accession: I49557

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-25 <RES>

A:Cross-references: GB:K01688; NID:G192246; PIDN:AAA37330.1; PID:G553881

R:Fenton, S.P.; Lanade, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.

Biochim. Biophys. Acta 1216, 469-474, 1993

A>Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.

A:Reference number: S39789; MUID:94092741; PMID:8268229

A:Accession: S39789

A:Molecule type: DNA

A:Residues: 1-80, 'E', 82-105, 'D', 107-185; 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-

R:Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Braindi, M.

Mol. Cell. Biol. 14, 5950-5960, 1994

A>Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indire

A:Reference number: I48300; MUID:94344105; PMID:8065328

A:Accession: I48300

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REP>

A:Cross-references: EMBL:X54876; NID:G50486; PIDN:CAA38657.1; PID:G50487

C:Genetics:

A:Gene: COL1A1

A:Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology

C:Keywords: coiled coil, extracellular matrix; glycoprotein; heterotrimer; triple helix

F; 1-22/Domain: signal sequence #status predicted <SIG>

F; 23-151/Domain: amino-terminal propeptide #status predicted <PRO>

F; 30-89/Domain: von Willebrand factor type C repeat homology <VWC>

F; 152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>

F; 1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match

Best Local Similarity

Matches 432; Conservative

22; Mismatches 162; Indels 117; Gaps

11;

```

QY 1 GPP-----GEPGPTGLPDPGERGGPSRGRFPAGADGVAGKPGPAGERGSGPPA 48
  |||
  |||
  |||
Db 447 GPPGPAGEGKRGARGEPGSLGPPGPGGSGRGRFPAGADGVAGKPGSGERGAGPPA 506

QY 49 GPKSGPAGGPGGAGLPGAKGLTGSPPSPDPDKTTPPPAGDGRPPGPPGARGQA 108
  |||
  |||
  |||
Db 507 GPKSGPAGGPGGAGLPGAKGLTGSPPSPDPDKTTPPPAGDGRPPGPPGARGQA 566

QY 109 GVMGPPGKGAAGBPKAGRGVPCPGAVGPPACKDGEAGAQGPPGPPAGPAGERGEGCPA 168
  |||
  |||
  |||
Db 567 GVMGPPGKGTAGEPGKAGRGVPCPGAVGPPACKDGEAGAQGPPGPPAGPAGERGEGCPA 626

QY 169 GSPGQGLPAGPAGERGKGEQGVPGDLGAPGSPGAGE-----PGP----- 212
  |||
  |||
  |||
Db 627 GSPGQGLPAGPAGERGKGEQGVPGDLGAPGSPGAGE-----PGP----- 686

QY 213 -----TGLPDPGERGGPSRGRFPAGADGVAGKPGPAGERGSGPPGKSP 258
  |||
  |||
  |||
Db 687 GNNGAPGNDGAKGDTGAFGAPGSGAQGLQMPGREGAAGLPGKGRDGDAGPKGADGSP 746

QY 259 GE-----AGRPGEAGLPGAKGLTGSPPSPDP-----GKTGPPGAGDGRP 300
  |||
  |||
  |||
Db 747 CKDGARGLTGPIGPPGAGAPGDKGEAGSPGPPGPTGARGAPGDRGBAGPPGAGFAGPP 806

QY 301 GPPGPPGARGQ-----AGVMGPPGPKGAAGEPGKAGRGVPGPPGA----- 341
  |||
  |||
  |||
Db 807 GADQOPGAKGPPGTGVKGDAGPPGAPGAPGPPGPIGNVAPGPKGPRGAAGPPGATGFP 866

QY 342 -----VGPAGKDGAGAQGPPGAPG-----AGERGEGQAGSPGQGLPAGPAGGAGKP 393
  |||
  |||
  |||
Db 867 GAAGKRGVPPGSGNAGPPGPPGVGKEGKGRGCTGTPAGRPGEVGPFGPPGAGEKGP 926

QY 394 GEQGVPGDLGNAPSGSAGPBEPTGLPDPGERGGPSRGRFPAGADGVAGKPGPAGERGSP 453
  |||
  |||
  |||
Db 927 GADGAPGSPGPPGQGTAGQGVVGLPFGQGRGEPGLPFPSPGEPKQSGSGSGRGP 986

QY 454 GPAGP-----KSPGPAEPGEGAGLPGAKGLTGSPPSPDPDKTTPPPAGDGRPPGPP 510
  |||
  |||
  |||
Db 987 GPMGPPGLAGPPGSGREGSPGAEPSGRDGAFGAKGDRGTGAPGPPGAPGAPGPPV 1046

QY 511 GARQAGVMGPPGPKGAAGBPKAGRGVPGP-----PGAV 546
  |||
  |||
  |||
Db 1047 GPAGKNGDRGTGAPGAPGPIGAGARGPAGPQGPGRGDKGTGEGDRTGKHRSGLQ 1106

QY 547 GPAGKDGAEAGAQGPPGAPGAGERGEGQAGSP-----GFGQLPAGPAGPGEAGKPGQGV 603
  |||
  |||
  |||
Db 1107 GPPGSPSPGEGQSGSAGPAGPRGPPGSGSGKDGGLNGLPPIGPPGPRGTGDSGPA 1166

QY 604 GDLGAPGSPGAP 616
  |||
  |||
  |||
Db 1167 GPPGPPGPPGPPG 1179

```

## RESULT 3

```

CGCHIS
collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
C;Species: Gallus gallus (chicken)
C;Date: 12-Aug-1981 #sequence revision 06-Jul-1982 #text_change 31-Mar-2000
C;Accession: A90458; A90181; A02857
R;Higberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.
Biochemistry 21, 2048-2055, 1982
A;Title: Amino acid sequence of chick skin collagen alpha1(I)-C88 and the complete prima
A;Reference number: A90458; MUID:82231995; PMID:70933229
A;Accession: A90458
A;Molecule type: protein
A;Residues: 1-1036 <HIG>
A;Experimental source: skin
A;Note: this is the latest in a series of papers from these workers elucidating the sequ
R;Eyre, D.R.; Glimcher, W.J.
Biochem. Biophys. Res. Commun. 48, 720-726, 1972
A;Title: Evidence for a previously undetected sequence at the carboxyterminus of the alp
A;Reference number: A90181; MUID:72243016; PMID:5047697
A;Accession: A90181

```

A;Molecule type: protein

A;Residues: 1037-1042 <EVR>

A;Experimental source: skin

A;Note: residues 1037-1042 above correspond to the carboxyl end of the protein

C;Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some

C;Comment: Most of the prolines at the third position of the tripeptide repeating unit

C;Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in po

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyrrolidone carboxylic acid; trimer;

F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 61.7%; Score 2151.5; DB 1; Length 1042;

Best Local Similarity 57.8%; Pred. No. 2.4e-101;

Matches 424; Conservative 17; Mismatches 175; Indels 117; Gaps 10;

QY 1 GPP-----GEPGPTGLPDPGERGGPSRGRFPAGADGVAGKPGPAGERGSGPPA 48

Db 296 GPPGPAGEGKRGARGEPGSLGPPGPGGSGRGRFPAGADGVAGKPGSGERGAGPPA 355

QY 49 GPKSGPAGGPGGAGLPGAKGLTGSPPSPDPDKTTPPPAGDGRPPGPPGARGQA 108

Db 356 GPKSGPAGGPGGAGLPGAKGLTGSPPSPDPDKTTPPPAGDGRPPGPPGARGQA 415

QY 109 GVMGPPGKGAAGBPKAGRGVPCPGAVGPPACKDGEAGAQGPPGPPAGPAGERGEGCPA 168

Db 416 GVMGPPGKGAAGBPKAGRGVPCPGAVGPPACKDGEAGAQGPPGPPAGPAGERGEGCPA 475

QY 169 GSPGQGLPAGPAGERGKGEQGVPGDLGAPGSPGAGE-----PGP----- 212

Db 476 GAPGQGLPAGPAGERGKGEQGVPGDNAGAPGARGERGPPGPPGPPGPP 535

QY 213 -----TGLPDPGERGGPSRGRFPAGADGVAGKPGPAGERGSGPP 251

Db 536 GANAPGNDGAKGDTGAFGAPGSGAQGLQMPGREGAAGLPGKGRDGDAGPKGADGSP 595

QY 252 -----AGPKSGPAGRP-----GEAGLPGAKGLTGSPPSPDPDKTTPPPGPA----- 294

Db 596 GKDLGRLGTGPIGPPGAPAGDKGEAGPPGAPGTCARGAPGDRGEPGPPGAPGAP 655

QY 295 -----GQDRGPPGPPGARGQAQVMGPPGPKGAAGBPKAGRGVPG 336

Db 656 GADQOPGAKGDTGADGAKGDPGPPGAPGTCAPGAGZVGAPGPKGARGSAGPPGATGFP 715

QY 337 GPPGAVGAPGKDGAGAQGPPGAPGAGE-----RGEQGPAGSPGQGLPAGPAGPGEAGKP 393

Db 716 GAARVVPGPPSGNIGLPGPPGPAKKGSKGPRGTGAPRPEBPAGPAGPAGPAGPAG 775

QY 394 GEQGVPGDLGAPGSPGAPGPPGPTGLPDPGERGGPSRGRFPAGADGVAGKPGPAGERGSP 453

Db 776 GADGPIGAPGTPGQIAGQGVVGLPQGRGERGFPGLPSPGEPKQSGSASGERGPP 835

QY 454 GPAGP-----KSPGPAEPGEGAGLPGAKGLTGSPPSPDPDKTTPPPAGDGRPPGPP 510

Db 836 GPMGPPGLAGPGEAGREGAPGAEGRDGAAGKPKGDRGTGAPGPPGAPGAPGAP 895

QY 511 GARQAGVMGPPGPKGAAGBPKAGRGVPGPPGAVGAPGAKDGE-----AGAQ 558

Db 896 GPAGKNGDRGTGTGAPGAPGPPGAPGAPGAPGPPGDKGTGEGQCDRGMKHRGFSGLQ 955

QY 559 GPP-----GPAGPAGERGEGQAGSPGQGLPAGPAGPAGPAGPAGPAG 603

Db 956 GPPGPPGAPGEGQSPGASGAPGPPGSGAAGKDGGLNGLPPIGPPGPRGTGEGVP 1015

QY 604 GDLGAPGSPGAP 616

Db 1016 GPPGPPGPPGPPG 1028

## RESULT 4

CORTIS

collagen alpha 1(I) chain - rat (tentative sequence) (fragments)

C;Species: Rattus norvegicus (Norway rat)

C;Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 31-Mar-2000

[illegible]

```

Db      509 GPSGARGE-----RFFGERGVQPPGPPAGPRGN-----NGAPGBB 544
QY      466 GRPGEAGLFGAKLGTSGSPGPGPKTGPFGPAQDGRPPGPPGARGAGVGMFPGP 525
Db      545 GAKGDTGAPGAPGSGGAPGLZGMSGLZGPPGPGS---PGZGSPGASGAPGR---GPP 598
QY      526 CAAREPKACRGVPGPPGAVGAPGACGACAGCAQPPGPPGAPGRGQGAGSPGQGLP 585
Db      599 GSASPGKGBLGLPGIPGPPGPRGRTGBAGSPGPPGPPG-----P 640
QY      586 GPAGPP 591
Db      641 GPPGPP 646

RESULT 5
745467
collagen alpha 1(II) chain precursor [imported] - horse
N:Alternate names: type II collagen
C:Species: Equus caballus (domestic horse)
C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C:Accession: T45467
R:Richardson, D.W.; Dodge, G.R.
submitted to the EMBL Data Library, June 1996
A:Description: Cloning of equine type II collagen and modulation of its expression in eq
A:Reference number: 222977
A:Accession: T45467
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1418 <RIC>
A:Cross-references: UNIPROT:Q28396; EMBL:U62528; PIDN:AA805773.1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 55.7%; Score 1944; DB 2; Length 1418;
Best Local Similarity 52.1%; Pred. No. 7.3e-91;
Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY      1 GPPCEPPTGLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 60
Db      156 GNPEPEPEGVSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 215
QY      61 -----GEAGLFGAKLGTSP---GSPGP-----DKGTGPPGPA--- 90
Db      216 GVKCHRGVPLDGAKEGAGAPGVKGESGSPGNGSPGMPGRLPGRGRTGPPAGAAGAR 275
QY      91 GQGRPPPPPPGARGAGVGMFPGPKGAAGEPGKAGER-----GVPPG- 134
Db      276 GNDGQGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 335
QY      135 -----PGAAGVAGKDXGAGAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 177
Db      336 GAAGNCTDGIPLGKAGSAGAPGIAGAPFPFGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 395
QY      178 GPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 237
Db      396 GEQGPKEPAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 455
QY      238 GPKGPAGERSPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 297
Db      456 GPKGAPGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 515
QY      298 GRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 357
Db      516 GRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 575
QY      358 GPAGPAGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 414
Db      576 GPAGPAGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 635
QY      415 GPTGLP-----GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 447
Db      636 GERGSPGAQGLQARGLPGTGTGDPKPGASGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 695

```

```

QY      448 GERGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 489
Db      696 GDRGVDGKPEBAGKDGKGRGLTGPIGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 755
QY      490 KGTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 534
Db      756 GETGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 815
QY      535 GERGVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 582
Db      816 GARGAQQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 875
QY      583 GLFGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 616
Db      876 GLQGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 909

RESULT 6
CGHU6C
collagen alpha 1(II) chain precursor [validated] - human
N:Alternate names: procollagen alpha 1(II) chain
N:Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen
C:Species: Homo sapiens (man)
C>Date: 28-May-1986 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C:Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S63
7250; I37251; I37252; I37253; I37254; I55338; I55335; I61910
R:Ryan, M.C.; Sieraski, M.; Sandell, L.J.
Genomics 8, 41-48, 1990
A:Title: The human type II procollagen gene: identification of an additional protein-cod
A:Reference number: A38513; MUID:91184811; PMID:2081599
A:Accession: A38513
A:Molecule type: DNA
A:Residues: 1-103 <RYA>
A:Cross-references: UNIPROT:P02458; UNIPROT:Q14042; UNIPROT:Q16672; UNIPROT:Q12985; UNIP
R:Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
Nucleic Acids Res. 17, 9473, 1989
A:Title: Nucleotide sequence of the full length cDNA encoding for human type II procolla
A:Reference number: S06715; MUID:90067946; PMID:2587267
A:Accession: S06715
A:Molecule type: mRNA
A:Residues: 1-28, 'R', '99-1487 <SU2>
A:Cross-references: EMBL:X16468; NID:g29515; PIDN:CAA34488.1; PID:g29516
A>Note: alternative splice form 1
R:Vikkula, M.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.
Biochem. J. 285, 287-294, 1992
A:Title: Structural analysis of the regulatory elements of the type-II procollagen gene.
A:Reference number: S24270; MUID:92344585; PMID:1637314
A:Accession: S24270
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-28 <VKS>
A:Cross-references: EMBL:X58709; GB:S40537; NID:g35659
A>Note: this translation is not annotated in GenBank entry HSFCOBL, release 111.0
R:Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
Gene 44, 11-16, 1986
A:Title: Promoter region of the human pro-alpha-1(II)-collagen gene.
A:Reference number: A24828; MUID:87031574; PMID:3021582
A:Accession: A24828
A:Molecule type: DNA
A:Residues: 1-8, 'T', '10-28 <NUN>
A:Cross-references: GB:M25698; NID:g180872; PIDN:AA52051.1; PID:g553237
R:Baldwin, C.T.; Reginaldo, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
Biochem. J. 262, 521-528, 1989
A:Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(II)
A:Reference number: S06496; MUID:90026318; PMID:2803268
A:Accession: S06496
A:Molecule type: mRNA
A:Residues: 7-28, 'R', '99-157, 'P', '159-440, 'G', '442-456, 'E', '458-640, 'A', '642-831, 'PA', '834, 'F'
A:Cross-references: EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID:g30041
A>Note: alternative splice form 1
R:Ryan, M.C.; Sandell, L.J.
J. Biol. Chem. 265, 10334-10339, 1990

```

A>Title: Differential expression of a cysteine-rich domain in the amino-terminal propept  
A/Reference number: A35428; MUID:90285153; PMID:2355003  
A/Accession: A35428  
A/Status: not compared with conceptual translation  
A/Molecule type: mRNA  
A/Residues: 27-81, 'L', 83-103 <RYA2>  
A/Note: alternative splice form 2; splicing appears to be under developmental regulation  
R;Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, P.  
Genomics 4, 438-441, 1989  
A>Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf  
A/Reference number: A30147; MUID:89233138; PMID:2714801  
A/Accession: A30147  
A/Molecule type: DNA  
A/Residues: 104-157, 'P', 159-236 <SUM>  
A/Cross-references: GB:J03065; GB:M23660; GB:M25656; GB:M25730; GB:M32168; GB  
P;Aia-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6585-6588, 1990  
A>Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pri  
A/Reference number: A94227; MUID:90370826; PMID:1975693  
A/Accession: A33116  
A/Molecule type: DNA  
A/Residues: 171-172, 'C', 174-175 <ALA>  
A/Note: mutant sequence from a family with family with primary generalized osteoarthritis  
R;Diab, M.; Wu, J.J.; Eyre, D.R.  
Biochem. J. 314, 327-332, 1996  
A>Title: Collagen type IX from human cartilage: a structural profile of intermolecular c  
A/Reference number: S64673; MUID:96195147; PMID:8660302  
A/Accession: S64674  
A/Molecule type: protein  
A/Residues: 188-189, 'X', 191-195; 1224-1230, 'X', 1232-1236 <DIA>  
R;Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,  
Eur. J. Biochem. 234, 125-131, 1995  
A>Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil can  
A/Reference number: S63514; MUID:96096730; PMID:8529631  
A/Accession: S63514  
A/Molecule type: protein  
A/Residues: 243-251; 575-590; 756-763, 'X', 765-779 <FPA>  
R;Tiller, G.E.; Weis, M.A.; Polunbo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,  
Am. J. Hum. Genet. 56, 388-395, 1995  
A>Title: An RNA-splicing mutation (G-51VS20) in the type II collagen gene (COL2A1) in a  
A/Reference number: I38867; MUID:95150028; PMID:7847372  
A/Accession: I38867  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 440, 'G', 442-456, 'E', 458-480, 'P', 482-509 <TLLA>  
A/Cross-references: EMBL:U15195; NID:9557053; PIDN:AA560370.1; PID:G557054  
R;Ramirez, F.  
submitted to the EMBL Data Library, December 1988  
A/Reference number: S04892  
A/Accession: S04892  
A/Molecule type: mRNA  
A/Residues: 501-676, 'A', 678-783, 'A', 785-831, 'PA', 834, 'P', 836-1214 <RAM>  
A/Cross-references: EMBL:X13783; NID:G30037; PIDN:CAA32030.1; PID:G930050  
R;Vikkula, M.; Peltonen, L.  
FEBS Lett. 250, 171-174, 1989  
A>Title: Structural analyses of the polymorphic area in type II collagen gene.  
A/Reference number: S05000; MUID:89325561; PMID:2753125  
A/Accession: S05000  
A/Molecule type: DNA  
A/Residues: 630-640, 'A', 642-785 <VIK2>  
A/Cross-references: EMBL:X16158; NID:G29951; PIDN:CAA34278.1; PID:G1335018; PIDN:CAA3427  
PIDN:CAA34283.1; PID:G1335023; PIDN:CAA34284.1; PID:G1335024  
R;Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D  
J. Biol. Chem. 267, 22522-22526, 1992  
A>Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain pro  
A/Reference number: A44309; MUID:93054548; PMID:1429602  
A/Accession: A44309  
A/Status: nucleic acid sequence not shown; not compared with conceptual translation  
A/Molecule type: DNA; mRNA  
A/Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'O', 1038-1052, 'E', 1054-1068, 'T',  
A/Cross-references: GB:L00977; NID:G180812; PIDN:AB23914.1; PID:G358774  
A/Note: sequence extracted from NCBI Backbone (NCBP:117273); parts of this sequence we  
A/Note: this translation is not annotated and this publication is not cited in GenBank s

A/Note: mutant sequence associated with perinatal lethal hypochondrogenesis  
R;Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.  
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990  
A>Title: Random duplication within a type II collagen gene (COL2A1) exon in an individu  
A/Reference number: S16502; MUID:90251662; PMID:2339128  
A/Accession: S16502  
A/Molecule type: DNA  
A/Residues: 1164-1184, 'GPSKDGANGIPGP', 1185-1199 <TIL2>  
A/Cross-references: EMBL:M37126; NID:G180808; PIDN:AAA52037.1; PID:G180809  
A/Note: mutant sequence from a patient with spondyloepiphyseal dysplasia  
R;Chen, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosfeld, F.G.; Solomon, E.  
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985  
A>Title: Identification and characterization of the human type II collagen gene (COL2A1  
A/Reference number: A02858; MUID:85190534; PMID:3857598  
A/Accession: A02858  
A/Molecule type: DNA  
A/Residues: 1032-1056, 'N', 1058-1069, 'T', 1070-1487 <CHE>  
A/Cross-references: GB:J00116; NID:G180395; PIDN:AAA51997.1; PID:G180396  
R;Elina, K.; Vuorio, T.; Vuorio, E.  
Nucleic Acids Res. 15, 9499-9504, 1987  
A>Title: Determination of the single polyadenylation site of the human pro-alpha-1(II)  
A/Reference number: A27280; MUID:88067771; PMID:2825137  
A/Accession: A27280  
A/Molecule type: DNA; mRNA  
A/Residues: 1175-1487 <ELI>  
A/Cross-references: EMBL:X06268; NID:G30096; PIDN:CAA29604.1; PID:G30097  
A/Experimental source: fetal epiphyseal cartilage  
R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.  
Biochem. J. 237, 923-925, 1986  
A>Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.  
A/Reference number: A57033; MUID:87099927; PMID:3800925  
A/Accession: A57033  
A/Molecule type: protein  
A/Residues: 'XE', 1244-1246, 'N', 1248, 'X', 1250-1265; 1295-1305; 1395-1408 <VAN>  
A/Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal pro  
R;Strom, C.M.; Upholt, W.B.  
Nucleic Acids Res. 12, 1025-1038, 1984  
A>Title: Isolation and characterization of genomic clones corresponding to the human ty  
A/Reference number: A21733; MUID:84118798; PMID:6320112  
A/Accession: A21733  
A/Molecule type: DNA  
A/Residues: 1245-1295 <STR1>  
A/Cross-references: EMBL:X00339; EMBL:X00298; NID:G394599; PIDN:CAA25092.1; PID:G437897  
A/Accession: B21733  
A/Molecule type: DNA  
A/Residues: 894-909, 'PE', <STR2>  
A/Cross-references: GB:K01785; NID:G30035; PIDN:CAA5082.1; PID:G1335032  
R;Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.  
Biochemistry 24, 6343-6348, 1985  
A>Title: Isolation and partial characterization of genomic clones coding for a human pr  
gene  
A/Reference number: A24561; MUID:86104139; PMID:3002437  
A/Accession: A24561  
A/Molecule type: DNA  
A/Residues: 1296-1358 <NUN2>  
A/Cross-references: GB:M12048; NID:G180017  
A/Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0  
A/Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with th  
R;Sangiorci, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez  
Nucleic Acids Res. 13, 2207-2225, 1985  
A>Title: Isolation and partial characterization of the entire human pro alpha 1(II) col  
A/Reference number: I37249; MUID:85215609; PMID:2987845  
A/Accession: S59491  
A/Molecule type: DNA  
A/Residues: 7-28, 'R', 99-114; 541-578; 786-802; 1055-1056, 'N', 1058-1068, 'T', 1070-1109; 1200-  
A/Accession: I84453  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 7-28 <GAN2>  
A/Cross-references: GB:M23759; NID:G180845; EMBL:X03320; GB:M24938; NID:G30104  
A/Note: the GenBank PID is based on an incorrect reading frame  
A/Accession: I37250  
A/Status: translated from GB/EMBL/DBJ



A:Molecule type: DNA

A:Residues: 541-560 &lt;SANK&gt;

A:Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621

A:Accession: I37251

Query Match 55.7%; Score 1943; DB 1; Length 1487;

Best Local Similarity 52.1%; Pred. No. 8.5e-91;

Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGEPGPTGLPGPPGGRGSGFFGADGVAGPKPAGERSPPGAGPKGSGFAGARP 60

DB 225 GNPEGPEGPGVSGPMGRGPPGPKGFGDDGAGKPKGAKGERGPPGQAGARFPGTGLP 284

QY 61 -----GEAGLPAGKGLTGSP---GSPGP-----DGKTPPGPA--- 90

DB 285 GVKHGRGYPGLDGAKEAGAPGVKESGSGPENGSPGMGRGLPGERGETGAGAGAR 344

QY 91 QDGRPPGPPGARGAQGVWGF-----GPKGAAGEPKKAGERGVPPGPGVGA 141

DB 345 GNDQGPAGPPGPPGAGGPPGFCAPCAKGEAGTTCARGPEGAQPRGPPGPGSPA 404

QY 142 -----GKDGEAGAQGPPGAGPAGERGEOGPAAGSGFQGLPGAPGPGGAGKPGGQVP 195

DB 405 GASNPGTGTLPGAKGAGAPGTAGAPFPGRPPDPQATGLPKGTGAPGTAGPK 464

QY 196 GD---LGAPGSPGAPGPTGLP-----GPPGERGPGSRGPPGADGVA 237

DB 465 GEQGPKEGPPGAGPQAGPPGAGEGKRGARPEGVGPIGPPGERGAPNRRGPPGQGLA 524

QY 238 GPKPAGERGSPGAGPKGSPGAGRPEAGLPGKAGLTGSPGSPDPDKTGPFGAG 297

DB 525 GPKGAPGERGSPGAGPKGANGDPGRPEGLPCARGLTGRPDAGPQGVKGSAGPAG 584

QY 298 GRPFPFPPGARGAQGVWGFPGKGAAGEPKKAGERGVPPGAVGPPAGKDGAGAQGPP 357

DB 585 GRPFPFPPGARGAQGVWGFPGKGAAGEPKKAGERGLPCAPGLRGLPKDGTGAE 644

QY 358 GPAGPAGERGSPGAGSGFQGLPGAPGPGGAGKPGQGVPPGDLGAPSGPAGE 413

DB 645 GPAGPAGERGSPGAGSGFQGLPGAPGPGGAGKPGQGVPPGDLGAPSGPAGE 704

QY 414 -----PGTGLP-----GPPGERGPGSRGPPGADGVA 447

DB 705 GERGSPGAQGLQGRGLPGTGTCTDGPCKASGPPAGPQAQGLQMPGERGAGTAGPK 764

QY 448 GERGS-----PGPAGPKSGPGEAGRPGEAGLPKAGLTGSP 483

DB 765 GDRGVDGKGPAGAPGKDGGRGLTGTPTGPPGPAAGKGEVPPGPAAGARGAPGR 824

QY 484 GSPGPDGKTGPPGAGDGRGPPGPPGARGAQGVWGFPGPKGAAGEP-----GKA 534

DB 825 GETGPTGSLGAPGADGQGKAGEGAGQGDAGAPGQPSGAPGQGTGVTGPK 884

QY 535 GERGVPPPGA-----VGPAGKDGAGAGQPPGPA---GPAGERGQGPAGSGFG 582

DB 885 GARGAQPPGATGPPGAAGRVGPPGNGNPPGPPGSKDGPCKGARGDSGPPGRAGEP 944

QY 583 GLCPGAPGPEAGKPGQGVPPGDLGARGPSGPAG 616

DB 945 GLQGPAGPKEGPEGDDGSGAEGPPGPOGLAG 978

RESULT 7

A41182

collagen alpha 1(II) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 13-Aug-1999

C:Accession: A41182; A44885

R:Metaseranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

J. Biol. Chem. 266, 16862-16869, 1991

A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and

A:Reference number: A41182; MUID:91358489; PMID:1865613

A:Accession: A41182

A&gt;Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-1419 &lt;MET&gt;

A:Cross-references: GB:M65161

R:Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.

Development 111, 945-953, 1991

A:Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilag

A:Reference number: A44885; MUID:91347939; PMID:1879363

A:Accession: A44885

A:Molecule type: DNA

A:Residues: 1-28 &lt;CHE&gt;

A:Cross-references: GB:S63180; NID:G234368; PIDN:AAB19627.1; PID:G234369

A&gt;Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBP:63192)

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime

F1191-1419/Domain: fibrillar collagen carboxyl-terminal homology &lt;FCC&gt;

Query Match 55.4%; Score 1932; DB 2; Length 1419;

Best Local Similarity 51.9%; Pred. No. 2.9e-90;

Matches 391; Conservative 36; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGEPGPTGLPGPPGGRGSGFFGADGVAGPKPAGERSPPGAGPKGSGFAGARP 60

DB 157 GNPEGPEGPGVSGPMGRGPPGPKGFGDDGAGKPKGAKGERGLPQAGARFPGTGLP 216

QY 61 -----GEAGLPAGKGLTGSP---GSPGP-----DGKTPPGPA--- 90

DB 217 GVKHGRGYPGLDGAKEAGAPGVKESGSGPENGSPGMGRGLPGERGTGAGAGAR 276

QY 91 QDGRPPGPPGARGAQGVWGF-----GPKGAAGEPKKAGERGVPPGPGVGA 141

DB 277 GNDQGPAGPPGPPGAGGPPGFCAPCAKGEAGTTCARGPEGAQPRGPPGPGSPA 336

QY 142 GKDE-----AGAQQPPGPPGAPGAGEGQGP-----AGSPGFGGLP 177

DB 337 GASNPCTGTGPAKGSAGAPGAGPFGTGPGRGPPGQATGTLGPKGQAGEGPIAGFK 396

QY 178 GPAGPAGERGSPGAGSGFQGLPGAPGPGGAGKPGQGVPPGDLGAPSGPAGE 237

DB 397 GDQGPKEGTPGAPGQAGPPGAGEGKRGARPEGAGPIGPPGERGAPNRRGPPGQGLA 456

QY 238 GPKPAGERGSPGAGPKGSPGAGRPEAGLPGKAGLTGSPGSPGDKTGPFGAG 297

DB 457 GPKAPAGERGSPGAGPKGANGDPGRPEGLPCARGLTGRPDAGPQGVKGSAGPAG 516

QY 298 GRPFPFPPGARGAQGVWGFPGKGAAGEPKKAGERGVPPGAVGPPAGKDGAGAQGPP 357

DB 517 GRPFPFPPGARGAQGVWGFPGKGAAGEPKKAGERGLPCAPGLRGLPKDGTGAG 576

QY 358 GPAGPAGERGSPGAGSGFQGLPGAPGPGGAGKPGQGVPPGDLGAPSGPAGE 413

DB 577 GPSGAGERGSPGAGSGFQGLPGAPGPGGAGKPGQGVPPGDLGAPSGPAGE 636

QY 414 -----PGTGLP-----GPPGERGPGSRGPPGADGVA 447

DB 637 GERGSPGAQGLQGRGLPGTGTCTDGPCKAAGPDPGPAQGLQMPGERGAGTAGPK 696

QY 448 GERGSPGAPKGSFGE-----AGRPGEAGLPKAGLTGSPGSGPAG 489

DB 697 GDRGVDGKGPAGAPGKDGGRGLTGTPTGPPGPAAGKGEVPPGSGTSGARGAPGE 756

QY 490 KTGTPPPAGQDGRPPGPPPGAR-----GQAGVMGFPFGPKGAAGEP-----GKA 534

DB 757 GETGPPGAPGAGPPGADGQPGAKGDGAGAPGQPSGAPGQGTGVTGPK 816

QY 535 GERGVPPPGA-----VGPAGKDGAGAGQPPGPA---GPAGERGQGPAGSGFG 582

DB 817 GARGAQPPGATGPPGAAGRVGPPGNGNPPGPPGSKDGPCKGARGDSGPPGRAGDP 876

QY 583 GLCPGAPGPEAGKPGQGVPPGDLGARGPSGPAG 616

DB 877 GLQGPAGPKEGPEGDDGSGAEGPPGPOGLAG 910







A:Reference number: A40333; MUID:92011890; PMID:1918153

A:Accession: A40333

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1492 <SUA>

A:Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M3596

A:Note: this sequence is presented as substitution relative to another sequence in a file es they replace; the appropriate interpretation of the sequence figure was reconstructed C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix F:37,96/Domain: von Willebrand factor type C repeat homology <WVC> F:1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 54.0%; Score 1883; DB 2; Length 1492;

Best Local Similarity 49.1%; Pred. No. 8.6e-88;

Matches 379; Conservative 46; Mismatches 185; Indels 162; Gaps 13;

QY 1 GPPPEPPTGLPGPGERGPGSGRPGCAD-----GVAGPKPAGERGSGPAGPK 51

DB 272 GPPQAGQAGPPTGPGVKGHGYPLDGGKGEAGAAKAGESGAGAPGMPGR 331

QY 52 GSPGEARPGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGAGVM 111

DB 332 GLPGERGRPGSSGAAGAR---GNDGLPGP---AGPPGVGPAAGPFGAPGSKGEAGPT 385

QY 112 GFPQPKAAGEPGKAGRGVPGP-----PCAVGPKDGEAGAQGPPGAGPA 159

DB 386 GARGPEGAQGRGSGTTPGSPGSGASGNPTDIPGAKGSSGASGTAGAPGPPGRGPP 445

QY 160 GEREQGP-----AGSPFGGLPCPAGPCEAGKPCGQVPGDLGAPGSGPAGBPPT 213

DB 446 GPQATGPLGPKGTGTPGTAGTGEHGPKEIGSAGPQAPGAGEGKRGARGPGAA 505

QY 214 GLPQPPGERGPGSRGPPGADGVAGPKPAGERGSPGAPKPSGPPGAGRPAGLPKAK 273

DB 506 GPLPFPGERGAPGNKPPQDGLAGPKGAPGERGVPLGGPKGNGDPGPPGPGPLPGAR 565

QY 274 GLTCSQSPGPDGKTGTPPAGODGRPGPPQARCOAGVMGPPGKGAAGPPGKAGER 333

DB 566 GLTGRPDAGQGVKVGSGAAGEDRGPPGPQARGQPVGMGPPGKANGPPGKAGEK 625

QY 334 GVPQPPGAVGPAKDGEAGAQGPPGAPAGERGEQGPAGSPGFGQLPGPAGPPGAGKP 393

DB 626 GLGAPGLRGLPGKDGETAQGNPAGPAGERGEQGPFGSPGQGLPGPPGEGGKP 685

QY 394 GEQVPGDL-----GAPPSGPA 411

DB 686 GDQVPEBAGAAGLVGPRGERGTFBERGSGPQGLQSGRLGPTGTPGPKGATGTPSGPN 745

QY 412 GEPGTLGPPPPBERG-----GP-----GSRGFPAGDVAGPKGPA 447

DB 746 CAQGPPLQGNPFGERGANGISGPKDRGDTGEKGPAGKDGSRGLTGLPGLPPGAPGN 805

QY 448 GERSGPGPAGP-----KGSFGRAGRPGA 471

DB 806 CEKGESGSPGPIVGARGAPGDRGNGPPGPAFAGPPGSDGQAGLKGDQSGSQKGA 865

QY 472 GLPKAKGLTSGSPGPDGKTGP-----PGPAGODGRPGP---PGPPGAR 513

DB 866 GAPQPGSGAPPGQPTGVNVPKGARGAQGAPAGATFPGAAGRVGTPGNNGPPGPP 925

QY 514 QGAVVMGPPGKGAAGEPKAGERGVFPGPAGVAGPKDGEAGAQGPPGAPAGERGEQ 573

DB 926 GSAGKEGKGVRGDAGPTGRAGDPLQGPAGAPGEKGEPEGSDGSPGDPGSPGQGLSGNR 985

QY 574 GPAGSP-----GQGLPCPAGPGEAGKP---GEQVPGDLGAPGSGGAG 616

DB 986 GIVGLPGQGRGRTFGLPSPGEPKQGGPGSGDRGPPGVPVPGTGPAG 1037

RESULT 12

S59856

collagen alpha 1(III) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C>Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text\_change 09-Jul-2004

C:Accession: S59856; S62120; S16373

R:Toman, P.D.; de Crombrughe, B.

Gene 147, 161-168, 1994

A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA

A:Reference number: S59856; MUID:95011609; PMID:7926795

A:Accession: S59856

A:Molecule type: DNA

A:Residues: 1-1464 <TOM>

A:Cross-references: UNIPROT:P08121; EMBL:X52046

R:Toman, D.

submitted to the EMBL Data Library, November 1994

A:Reference number: S62120

A:Accession: S62120

A:Molecule type: DNA

A:Residues: 1-866, 'G', 868-1464 <TOA>

A:Cross-references: EMBL:X52046; NID:9575321; PIDN:CAA36279.1; PID:9575322

R:Yetsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

Biochim. Biophys. Acta 1089, 241-249, 1991

A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.

A:Reference number: S16176; MUID:91274355; PMID:2054384

A:Accession: S16373

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1442-1464 <MET>

A:Cross-references: EMBL:X57993; NID:950476; PIDN:CAA41048.1; PID:950477

C:Genetics:

A:Introns: 23/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 2

58/3; 673/3; 706/3; 742/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology

C:Keywords: coiled coil; extracellular matrix

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-154/Domain: propetide #status predicted <PRO>

F:32-92/Domain: von Willebrand factor type C repeat homology <WVC>

F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>

F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 53.3%; Score 1857.5; DB 2; Length 1464;

Best Local Similarity 53.4%; Pred. No. 1.6e-86;

Matches 369; Conservative 39; Mismatches 190; Indels 93; Gaps 12;

QY 1 GPPPEPPTGLPGPGERGPGSGRPGCADGVAGPKPAGERGSGPAGPKSGPGEAGRP 60

DB 467 GSPGPEGANGLPGAAGRGPSGFRGPAFNGIPGKPPGERGGPGPAGPRGAVGEPGRD 526

QY 61 GSAGLPGAKGLTSGSPGPDGKTGPPGAGDGRPPGPPGARGQAGVMGPPGPKGNA 120

DB 527 GTFPGGIRGMPGSPGPGNDKPGPPGSGSGSPGPPGSGRPGQPGVMGPPGPKGND 566

QY 121 GPPKAGERGVPPGAVGPAKDGAGAQGPPGAPAGERGEQGPAGSGPFGQLPGPA 180

DB 587 GAPKNGERGPGGPGPLPGPAGKNGETGPGPPGTGTAGDKGDSGPPGPGQLGIPGTG 646

QY 181 GPPGEAGKPGEGVPGDGLGAPGSPGAPGEPGTGLPGPPGERGGPGSGRPGFAGDVAGPK 240

DB 647 GPPGNGKPGEPGPKGEVGP-----GAPGKGDSGAPGERGPGTGTAGIPARGAGGPP 700

QY 241 GPAGERGSPGAPGKSGPGEAGRPGEAGLPGAKGLTSGSPGPGDKTGTGPPGAGDGRP 300

DB 701 GP---EGGKGPAGPPGPPGASGSLQGNPGR---GGPGSPCKGKEGEPGAGADVP 754

QY 301 GPPGPPGARGQAGVMGPPGPKGAAGEPGKAGRGVPGPAGVGPAGKDGAEAGAQGPPGPA 360

DB 755 GKDGP---RGAPGPIPPGP---AGQPGDKGSGSPGLPGIAGPRGPGGERGHPGPPGA 808

QY 361 -----GPAGERGEQGPAGSPFGQLPGPAGPGE---AGKPGECQVPGDLGAPGSP 408

DB 809 GPPGAPQNGEPGKAGERGAPGKGGGPPGAGTGTGSSGPPGPPGPPGQGVKSGRSGPP 868

QY 409 GPAGEBPPTGLPQPPGERGGPSRGPFGADGVAGPKPAGERGSGPAGPKSGPGEAGRP 468

DB 869 GTAGFPGGRLPGPPGNNGNPPGPPSPGAPKDGPPGAGNSGSPGNPPIAGPKGDAGQP 928

QY 469 CE-----AGLPGAKGLTSGPSGPDGKTGP-----PGPAGQD 501  
 Db 929 CEKPPGAGQPPGSPGLIAGTLGARGLAGPFGMPGPRGSPGQGIKGSXGKPGASGHN 988  
 QY 502 GRPPPPG-----PGARQAGVMGFPFGPKGAAGP---PGKAGRGVP 540  
 Db 989 GERFPFGQGLPGQGTAGEPRDNGSDQPGDRSGPGKGDGNGSGFGAPGAPHP 1048  
 QY 541 GPPGAVGAGKXGFE-----AGAGPPGPGAPGAGRGQGPAGSPFGQGLP 585  
 Db 1049 GPPGVPVPSGSGRGTGTPAGPGACGAPGAPGQPRGDKGTGTRGSGNGIKGHR 1108  
 QY 586 GPAGPPGAGKPGGQVPGDLGAPSGPAG 616  
 Db 1109 GPPGPPGPPGAGGQAGTSGSPGAPRG 1139

RESULT 13  
 CGB07S  
 collagen alpha 1(III) chain - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 09-Jul-2004  
 C:Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946  
 R:Fietzek, P.P.; Allmann, H.; Rauterberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979  
 A:Title: The covalent structure of calf skin type III collagen. I. The amino acid sequence  
 A:Reference number: A02862; PMID:80026026; PMID:488906  
 A:Accession: A02862  
 A:Molecule type: protein  
 A:Residues: 1-242 <FIB>  
 A:Cross-references: UNIPROT:P04258  
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979  
 A:Title: The covalent structure of calf skin type III collagen. II. The amino acid sequence  
 A:Reference number: A38001; PMID:80026027; PMID:488907  
 A:Accession: A38001  
 A:Molecule type: protein  
 A:Residues: 243-422 <DEWI>  
 R:Bentz, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979  
 A:Title: The covalent structure of calf skin type III collagen. III. The amino acid sequence  
 A:Reference number: A38002; PMID:80026028; PMID:488908  
 A:Accession: A38002  
 A:Molecule type: protein  
 A:Residues: 423-571 <BEN>  
 R:Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979  
 A:Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequence  
 A:Reference number: A38003; PMID:80026029; PMID:488909  
 A:Accession: A38003  
 A:Molecule type: protein  
 A:Residues: 572-808 <LAN>  
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979  
 A:Title: The covalent structure of calf skin type III collagen. V. The amino acid sequence  
 A:Reference number: A38004; PMID:80026030; PMID:488910  
 A:Accession: A38004  
 A:Molecule type: protein  
 A:Residues: 809-947 <DEW2>  
 R:Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979  
 A:Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequence  
 A:Reference number: A38005; PMID:80026031; PMID:488911  
 A:Accession: A38005  
 A:Molecule type: protein  
 A:Residues: 948-1049 <ALL>  
 A:Experimental source: skin  
 R:Henkel, W.  
 Biochem. J. 318, 497-503, 1996  
 A:Title: Cross-link analysis of the C-telopeptide domain from type III collagen.  
 A:Reference number: S71946; PMID:86404897; PMID:8809038  
 A:Accession: S71946

A:Molecule type: protein  
 A:Residues: 87-1061/1017-1029,1037-1049 <HEN>  
 C:Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are highly conserved. The type III collagen molecule is a trimer of identical chains, linked to each other by interchain disulfide bonds. The type III collagen molecule is a trimer of identical chains, linked to each other by interchain disulfide bonds. The type III collagen molecule is a trimer of identical chains, linked to each other by interchain disulfide bonds.  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline; F:1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>  
 F:1-14/Region: amino-terminal nonhelical telopeptide  
 F:15-1040/Region: helical  
 F:587-589/Region: cell attachment (R-G-D) motif  
 F:752-754/Region: cell attachment (R-G-D) motif  
 F:875-877/Region: cell attachment (R-G-D) motif  
 F:878-880/Region: cell attachment (R-G-D) motif  
 F:935-937/Region: cell attachment (R-G-D) motif  
 F:1041-1049/Region: carboxyl-terminal nonhelical telopeptide  
 F:95,107,119,938,950/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:107,950/Modified site: allylsine (Lys) #status predicted  
 F:107/Binding site: carbohydrate (Lys) (covalent) #status experimental  
 F:1040,1041/Disulfide bonds: interchain #status predicted

Query Match 53.2%; Score 1856; DB 1; Length 1049;  
 Best Local Similarity 51.2%; Pred. No. 1.5e-86;  
 Matches 371; Conservative 34; Mismatches 211; Indels 108; Gaps 14;  
 QY 1 GPPGEPGPTGLPFGPGRGSGRFGPGADGVAGPKGFPAGRGSGFPAGPKSGPGEAGRP 60  
 Db 312 GSPGEPGANGLPGAAGRGVFGFRGPGAGANGLPCEKGGPPGDRGGPGAGPRGVAGEPGRN 371  
 QY 61 GEAGLPKAGLGTSGPSGPDGKTGPPGAGQDGRGPPGPPGARGQAGVMGFPKGA 120  
 Db 372 GLPGGGLRGIFGSPGSGSGKTPPGSQGETGRPPGSPGPRGPGVGMGFPKGGND 431  
 QY 121 CEPKAGRGVPPGPGAVGPGAGKXGFEAGAGQPPGAPGAGRGQGPAGSPFGQGLPGPA 180  
 Db 432 GAFNGRGGFGPGPGPGAGKNGETGPPQPPPTGPGDKGTGTPGQGLGSGTS 491  
 QY 181 GPPGAGKPGQGVPGDLGAP-----GPSGAPGEPGTGLPFPGRGPGSGRFFGAD 234  
 Db 492 GPPGNGKPGFPGPKGEAGAPGIPGKGDGSGAPGERGPPGAGGPPGPRGAGPPGPGSGK 551  
 QY 235 GVAGPKGAPGERSP---GPAGPKGSGCEAGRPGCEAGLPGAKGLTSGSPGPDGKT--- 288  
 Db 552 GAAGPFGPGSAGTGLGMPGRRGGPGGPGPKDKKEPSSGVDDGAFGKDGPRGTGPI 611  
 QY 289 GPPGAGQDGR-----PQPPGPPGARGQ---AGVMGPPGPKGAAGEFGKAGER 333  
 Db 612 GPPGAGQPDGKSGSAPGVPIAGPRGGPGERGEGQPPGAPGAPGQNGEPGAKGER 671  
 QY 334 GVP-----GPPGAVGPRAGKXGFEAGAGQPP-----GPAGPAGER 366  
 Db 672 GAPGKGGGPPGAAGPAGSGGPPGPPGQVKGERSGPGGPAAGPFGGRGPPGPGSN 731  
 QY 367 GEQGPAGSPGQGLPGPAGPPEAGKXGFEQGV-----PGDLGAPGSGPAGEPGPT 417  
 Db 732 GNPFGPGSSGAPGKDGPPGPGSNGAFSGPGISGPKGDSGPPGERGAPGPGPPGAPGL 791  
 QY 418 GL-----PQPPGERGGSGRFGPGADGVAGPKGFPAGRGSGRFP-----AGPK 459  
 Db 792 GIAGLTGARGLAGPFGMPGARGSGFPQGIKGENKPKPGPSQNGRGPFGQGLPGLAGTA 851  
 QY 460 GSPGAGRPFAGLPGA-----KGLTSGPSGPDGKTGPPGP-----AGQDGRP 504  
 Db 852 GEPGPDGNGSDGLPGRDAGPAKXDRGNGSPGAPGAPGHPGPPGPPGVPAGKSGDRGET 911  
 QY 505 GPPGPPGARGQAGVMGFPFGPKGAAGEFGKAGERGV-----PQPPGAVGAPGKXGFE 555  
 Db 912 GPAGPSGAPGAPGSRGPPGPGQPRDKGTGTERGAMGIKGHRGPPGPPGAPGPPGAGH 971  
 QY 556 GAQGPFGPAGPAGRGSGQGPAGSPGFOGLPGAPGPPGCEAGKPGP-----QGVPGDLGAPG 612  
 Db 972 GAVSGPAGPAGPAGPAGSGPPGPKDGASGHFGPTGPPGPRGNRGERSGSGFPGHQGP 1031  
 QY 613 GPAG 616

Db 1032 GPPG 1035

RESULT 14

CGHUTL

collagen alpha 1(III) chain precursor - human

N:Alternate names: procollagen alpha 1(III) chain

C:Species: Homo sapiens (man)

C:Date: 24-Apr-1984 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004

C:Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90399; Prockop, D.J.

R:Prockop, D.J.

submitted to the EMBL Data Library, February 1989

A:Reference number: S05272

A:Accession: S05272

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1240, 'V', 1242-1466 <PRC>

A:Cross-references: UNIPROT:P02461; ENBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058

R:Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.

Biochem. J. 260, 509-516, 1989

A:Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human

ences.

A:Reference number: S04642; MUID:89350838; PMID:2764886

A:Accession: S04642

A:Molecule type: mRNA

A:Residues: 1-1196 <ALA>

A:Cross-references: ENBL:X14420; NID:G30057; PIDN:CAA332583.1; PID:G30058

A>Note: the complete sequence is not shown

R:Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.

Gene 78, 255-265, 1989

A:Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene

A:Reference number: PE0011; MUID:89378752; PMID:2777083

A:Accession: PE0011

A:Molecule type: DNA

A:Residues: 1-176 <BEN>

A:Cross-references: GB:M26939; NID:G180813; PIDN:AA52040.1; PID:G180814

R:Tomán, P.D.; Ricca, G.A.; de Crombrughe, B.

Nucleic Acids Res. 16, 7201, 1988

A:Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre

A:Reference number: S01726; MUID:88303360; PMID:3405773

A:Accession: S01726

A:Molecule type: mRNA

A:Residues: 1-170 <TOM>

A:Cross-references: ENBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:G30061

A>Note: the authors translated the codon CAG for residue 154 as His

R:Janeczko, R.A.; Ramirez, F.

Nucleic Acids Res. 17, 6742, 1989

A:Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.

A:Reference number: S04887; MUID:89386015; PMID:2780304

A:Accession: S04887

A:Molecule type: mRNA

A:Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634, 'A';

A:Cross-references: ENBL:X15332; NID:G29545; PIDN:CAA33387.1; PID:G30045

A>Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide

R:Sever, J.M.; Kang, A.H.

Biochemistry 16, 1158-1164, 1977

A:Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide

A:Reference number: A90399; MUID:77134724; PMID:557335

A:Accession: A90399

A:Molecule type: protein

A:Residues: 'V', 159-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>

A:Experimental source: liver

A>Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact

R:Sever, J.M.

submitted to the Atlas, December 1977

A:Reference number: A94562

A:Accession: A94562

A:Molecule type: protein

A:Residues: 'V', 163-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>

A:Experimental source: liver

A>Note: author submitted corrections to A90399

R:Miliewicz, D.M.; Witz, A.M.; Manchester, D.K.; Waldstein, G.; Byers, P.H.

Am. J. Hum. Genet. 53, 62-70, 1993

A:Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual

spring.

A:Reference number: I51868; MUID:93304430; PMID:8317500

A:Accession: I51868

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 186-194 <MIL>

A:Cross-references: GB:S62925; NID:G386425; PIDN:AAD13937.1; PID:G4261637

R:Chiodo, A.A.; Silencio, D.O.; Cole, W.G.; Bateman, J.F.

Biochem. J. 311, 939-943, 1995

A:Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL

A:Reference number: S59511; MUID:96067614; PMID:7487954

A:Accession: S59511

A:Molecule type: mRNA

A:Residues: 302-423 <CHI>

A:Cross-references: GB:S79877; NID:gl195576; PIDN:AA35615.1; PID:gl195577

R:Sever, J.M.; Kang, A.H.

Biochemistry 17, 3404-3411, 1978

A:Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr p

A:Reference number: A90414; MUID:79000343; PMID:687591

A:Accession: A90414

A:Molecule type: protein

A:Residues: 399-675, 'N', 677-727 <SEY3>

A:Experimental source: liver

R:Lee, B.; Vitale, E.; Superi-Furga, A.; Steinmann, B.; Ramirez, F.

J. Biol. Chem. 266, 5256-5259, 1991

A:Title: G to T transversion at position +5 of a splice donor site causes skipping of t

A:Reference number: I55349; MUID:91161621; PMID:1672129

A:Accession: I55349

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 537-605 <LEE>

A:Cross-references: GB:M59312; NID:G180815; PIDN:AAA52041.1; PID:G180816

R:Sever, J.M.; Mainardi, C.; Kang, A.H.

Biochemistry 19, 1583-1589, 1980

A:Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from t

A:Reference number: A90438; MUID:80188282; PMID:6246925

A:Accession: A90438

A:Molecule type: protein

A:Residues: 728-895, 'A', 897-964 <SEY4>

A:Experimental source: liver

R:Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Cha

J. Biol. Chem. 265, 17070-17077, 1990

A:Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping a

A:Reference number: A38303; MUID:91009133; PMID:2145268

A:Accession: A38303

A:Molecule type: mRNA

A:Residues: 861-1015 <COL>

A:Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AA59383.1; PID:

A>Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos sy

R:Mankoo, B.S.; Dalgleish, R.

Nucleic Acids Res. 16, 2337, 1988

A:Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.

A:Reference number: S02119; MUID:88189827; PMID:3357782

A:Accession: S02119

A>Status: translation not shown

A:Molecule type: mRNA

A:Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>

A:Cross-references: ENBL:X06700; NID:G30053; PIDN:CAA29886.1; PID:G30054

R:Sever, J.M.; Kang, A.H.

Biochemistry 20, 2621-2627, 1981

A:Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from t

A:Reference number: A90446; MUID:81208139; PMID:7016180

A:Accession: A90446

A:Molecule type: protein

A:Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157

A:Experimental source: liver

R:Loide, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; My

Nucleic Acids Res. 12, 9383-9394, 1984

A:Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollag

A:Reference number: A93551; MUID:85087944; PMID:6096827

A:Accession: A93551



Search completed: October 18, 2004, 13:42:52  
Job time : 29.4569 secs

*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:23:10 ; Search time 110.271 Seconds  
(without alignments)  
3219.411 Million cell updates/sec

Title: US-10-658-989A-2

Perfect score: 3488

Sequence: 1 GPPGCPPTGLPQPPGRRG.....GEGVPGDLGAPGSPGAGG 617

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 57537466 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt\_02:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2237.5	64.1	1464	1	CALL_HUMAN
2	2233.5	64.0	1461	2	P02452
3	2233.5	64.0	1464	2	O76045
4	2231.5	64.0	1460	1	Q8N473
5	2222	63.7	1069	2	Q6LAN8
6	2222	63.7	1069	2	CAA67261
7	2215.5	63.5	1453	2	Q63079
8	2205.5	63.2	1453	1	CALL_MOUSE
9	2205.5	63.2	1453	2	Q61059
10	2176.5	62.4	1225	2	Q6PCL3
11	2176.5	62.4	1225	2	AAH59281
12	2151.5	61.7	1453	1	CALL_CHICK
13	2043.5	58.6	1450	2	Q9YIB4
14	2043.5	58.6	1445	2	Q93251
15	2029	58.2	1449	2	Q80285
16	2010.5	57.6	671	1	CALL_RAT
17	1976.5	56.7	1447	2	Q9IB91
18	1953	56.0	1160	2	Q14046
19	1953	56.0	1487	2	Q14047
20	1952	56.0	1487	2	O77753
21	1944	55.7	1418	2	Q28396
22	1943	55.7	1418	1	CALL_HUMAN
23	1942	55.7	1420	2	Q90W37
24	1937	55.5	1269	2	O7T227
25	1936	55.5	1442	2	Q62031
26	1936	55.5	1442	2	Q62031
27	1936	55.5	1459	1	CALL_MOUSE
28	1936	55.5	1459	2	Q62032
29	1927	55.2	1419	2	Q80VX3
30	1927	55.2	1419	2	Q80X38
31	1923	55.1	1447	2	Q6ULU5

32 1923 55.1 1447 2 AAR24536  
33 1921 55.1 1419 2 Q63123  
34 1914 54.9 1447 2 Q6P4U1  
35 1914 54.9 1447 2 AAR63249  
36 1898 54.4 779 1 CALL\_BOVIN  
37 1892 54.2 1492 2 Q6P4Z2  
38 1892 54.2 1492 2 AAR63191  
39 1890.5 54.2 1486 2 Q1717  
40 1888.5 54.1 1486 2 Q72TI6  
41 1883 54.0 1418 2 Q9W7R9  
42 1880 53.9 1491 2 Q7ZTM4  
43 1880 53.9 1491 2 Q1718  
44 1873 53.7 1262 1 CALL\_CHICK  
45 1873 53.7 1449 2 Q910C0

## ALIGNMENTS

RESULT 1  
CALL\_HUMAN  
ID CALL\_HUMAN STANDARD; PRT; 1464 AA.  
AC P02452; P78441; Q13896; Q13902; Q13903; Q14037; Q14992; Q15176;  
AC Q15201; Q16050; Q7K230; Q7K234; Q8IV15; Q9UML6; Q9UMW7;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Collagen alpha 1(I) chain precursor.  
GN Name=COL1A1;  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND VARIANT SER-1434.  
RA Dalglish R.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBSJ databases.  
RN [2]  
RP SEQUENCE OF 1-589 FROM N.A.  
RX PubMed=2843432;  
RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,  
RA Pretorius P.J.;  
RT "Complete nucleotide sequence of the region encompassing the first  
twenty-five exons of the human pro alpha 1(I) collagen gene  
(COL1A1).";  
RL Gene 67:103-115(1988).  
RN [3]  
RP SEQUENCE OF 1-472 FROM N.A.  
RX MEDLINE=89025644; PubMed=3178743;  
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,  
JAenisch R., Prockup D.J.;  
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain  
of human type I procollagen.";  
RL Biochem. J. 253:919-922(1988).  
RN [4]  
RP SEQUENCE OF 1-181 FROM N.A.  
RX MEDLINE=84720697; PubMed=6462220;  
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,  
Myers J., Williams C., Ramirez F.;  
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary  
conservation of a pattern of introns and exons.";  
RL Nature 310:337-340(1984).  
RN [5]  
RP SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.  
RC TISSUE=Skin;  
RX MEDLINE=71038625; PubMed=5529814;  
RA Click E.M., Bornstein P.;  
RT "Isolation and characterization of the cyanogen bromide peptides from  
the alpha 1 and alpha 2 chains of human skin collagen.";  
RL Biochemistry 9:4699-4706(1970).  
RN [6]  
RP SEQUENCE OF 425-1464 FROM N.A.  
RX MEDLINE=84080385; PubMed=6689127;

RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,  
 RA Prockop D.J.;  
 RA "Nucleotide sequences of complementary deoxyribonucleic acids for the  
 RT pro alpha 1 chain of human type I procollagen. Statistical evaluation  
 RT of structures that are conserved during evolution.";  
 RL Biochemistry 22:5213-5223(1983).  
 RN [7]  
 RN SEQUENCE OF 472-607 FROM N.A.  
 RP PubMed=2981843;  
 RX Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;  
 RA "Multixon deletion in an osteogenesis imperfecta variant with  
 RT increased type III collagen mRNA.";  
 RL J. Biol. Chem. 260:691-694(1985).  
 RN [8]  
 RN SEQUENCE OF 488-625 FROM N.A.  
 RP PubMed=3857621;  
 RX Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;  
 RA "Intron-mediated recombination may cause a deletion in an alpha 1 type  
 RT I collagen chain in a lethal form of osteogenesis imperfecta";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874(1985).  
 RN [9]  
 RN SEQUENCE OF 1179-1464 FROM N.A., VARIANTS OI-III H1S-1277; ARG-1388 AND  
 RP 1337-GUJ-NYR-1338 DEL. AND VARIANT SER-1434.  
 RX MEDLINE=93352646; PubMed=8349697;  
 RA Chessler S.D., Wallis G.A., Byers P.H.;  
 RT "Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)  
 RT chain of type I collagen result in defective chain association and  
 RT produce lethal osteogenesis imperfecta";  
 RL J. Biol. Chem. 268:18218-18225(1993).  
 RN [10]  
 RN SEQUENCE OF 1229-1454 FROM N.A.  
 RP TISSUE=bone;  
 RX MEDLINE=88124208; PubMed=3340531;  
 RA Maekelae J.K., Raassina M., Virta A., Vuorio E.;  
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide  
 RT domain";  
 RL Nucleic Acids Res. 16:349-349(1988).  
 RN [11]  
 RN SEQUENCE OF 1-34 FROM N.A.  
 RX MEDLINE=85130970; PubMed=2857713;  
 RA Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;  
 RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.  
 RT Promoter structure, AluI repeats, and polymorphic transcripts";  
 RL J. Biol. Chem. 260:2315-2320(1985).  
 RN [12]  
 RN SEQUENCE OF 1-34 FROM N.A.  
 RX MEDLINE=88097389; PubMed=3480516;  
 RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;  
 RT "Regulatory elements in the first intron contribute to transcriptional  
 RT control of the human alpha 1(I) collagen gene";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).  
 RN [13]  
 RN SEQUENCE OF 1-44 FROM N.A.  
 RP MEDLINE=88033098; PubMed=2822714;  
 RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,  
 RA de Wet W.;  
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)  
 RT collagen gene enhance transcription";  
 RL J. Biol. Chem. 262:15151-15157(1987).  
 RN [14]  
 RN SEQUENCE OF 175-187 AND 274-289.  
 RP PubMed=2169412;  
 RX Baetge B., Notbohm H., Diebold J., Lehmann H., Bodo M., Deutzmann R.,  
 RA Muller P.K.;  
 RT "A critical crosslink region in human-bone-derived collagen type I.  
 RT Specific cleavage site at residue Leu95.";  
 RL Eur. J. Biochem. 192:153-159(1990).  
 RN [15]  
 RN SEQUENCE OF 263-268.  
 RP TISSUE=skin;  
 RX MEDLINE=71001508; PubMed=4319110;  
 RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;  
 RT "A comparative study of glycopeptides derived from selected vertebrate

RT collagens. A possible role of the carbonyl group in fibril formation.";  
 RL J. Biol. Chem. 245:5042-5048(1970).  
 RN [16]  
 RN SEQUENCE OF 281-302; 402-420; 823-842; 924-944; 1026-1045 AND  
 RP 1143-1162 FROM N.A.  
 RX PubMed=2374517;  
 RA Labhard M.E., Hollister D.W.;  
 RT "Segmental amplification of the entire helical and telopeptide regions  
 RT of the cDNA for human alpha 1 (I) collagen.";  
 RL Matrix 10:124-130(1990).  
 RN [17]  
 RN SEQUENCE OF 425-490; 965-1024; 999-1039 AND 1453-1464 FROM N.A.  
 RP MEDLINE=83064528; PubMed=6183642;  
 RX Chu M.-L., Myers J.C., Bernard M.P., Ding J.-F., Ramirez F.;  
 RA "Cloning and characterization of five overlapping cDNAs specific for  
 RT the human pro alpha 1(I) collagen chain";  
 RL Nucleic Acids Res. 10:5925-5934(1982).  
 RN [18]  
 RN SEQUENCE OF 710-745 FROM N.A., AND VARIANT OI-II ARG-728.  
 RP PubMed=2339700;  
 RX Wallis G.A., Starman B.J., Zinn A.B., Byers P.H.;  
 RA "Variable expression of osteogenesis imperfecta in a nuclear family is  
 RT explained by somatic mosaicism for a lethal point mutation in the  
 RT alpha 1(I) gene (COL1A1) of type I collagen in a parent.";  
 RL Am. J. Hum. Genet. 46:1034-1040(1990).  
 RN [19]  
 RN SEQUENCE OF 746-781 FROM N.A., AND VARIANT OI-III SER-767.  
 RP MEDLINE=95187161; PubMed=7881420;  
 RX Forlino A., Zolezzi F., Valli M., Pignatti P.F., Cetta G.,  
 RA Brunelli P.C., Mettes M.;  
 RT "Severe (type III) osteogenesis imperfecta due to glycine  
 RT substitutions in the central domain of the collagen triple helix";  
 RL Hum. Mol. Genet. 3:2201-2206(1994).  
 RN [20]  
 RN SEQUENCE OF 1187-1220 FROM N.A., AND VARIANT CVS-1195.  
 RP PubMed=3170557;  
 RX Cohn D.H., Apone S., Eyre D.R., Starman B.J., Andressen P.,  
 RA Charbonneau H., Nicholls A.C., Pope F.M., Byers P.H.;  
 RT "Substitution of cysteine for glycine within the carboxyl-terminal  
 RT telopeptide of the alpha 1 chain of type I collagen produces mild  
 RT osteogenesis imperfecta";  
 RL J. Biol. Chem. 263:14605-14607(1988).  
 RN [21]  
 RN SEQUENCE OF 1440-1454 FROM N.A.  
 RP MEDLINE=90110490; PubMed=2295701;  
 RA Willing M.C., Cohn D.H., Byers P.H.;  
 RT "Frameshift mutation near the 3' end of the COL1A1 gene of type I  
 RT collagen predicts an elongated pro alpha 1(I) chain and results in  
 RT osteogenesis imperfecta type I";  
 RL J. Clin. Invest. 85:282-290(1990).  
 RN [22]  
 RN SEQUENCE OF 1454-1464 FROM N.A.  
 RP PubMed=1995349;  
 RX Maatta A., Bornstein P., Penttinen R.P.;  
 RA "Highly conserved sequences in the 3'-untranslated region of the  
 RT COL1A1 gene bind cell-specific nuclear proteins";  
 RL FEBS Lett. 279:9-13(1991).  
 RN [23]  
 RN REVIEW ON VARIANTS.  
 RX MEDLINE=91184577; PubMed=2010058;  
 RA Kuivaniemi H., Tromp G., Prockop D.J.;  
 RT "Mutations in collagen genes: causes of rare and some common diseases  
 RT in humans";  
 RL FASEB J. 5:2052-2060(1991).  
 RN [24]  
 RN REVIEW ON VARIANTS.  
 RX MEDLINE=97255959; PubMed=9101290;  
 RA Kuivaniemi H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 RT associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels";  
 RL Hum. Mutat. 9:300-315(1997).  
 RN [25]

```
RP REVIEW ON OI VARIANTS.
RX MEDLINE=91374476; PubMed=1895312;

Query Match      64.1%; Score 2237.5; DB 1; Length 1464;
Best Local Similarity 60.2%; Pred. No. 9.5e-83;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

Qy 1 GPP-----GPGGTGTLPGPPGRRGGSGRFPQADGAVAGPKGPPAGRGSPGPA 49
Db |||||
458 GPPGPAEGKRGARGSPGFTGLPGPPGRRGGSGRFPQADGAVAGPKGPPAGRGSPGPA 517
Qy 49 GPKGSGEAGRPAGLPGAKGLTGTSGSGSPGDKTGPAGDGRPPGPPGARGOA 108
Db |||||
518 GPKGSGEAGRPAGLPGAKGLTGTSGSGSPGDKTGPAGDGRPPGPPGARGOA 577
Qy 109 GVMGFPFGKGAAGFPKAGRGVPGPPGAVGPAKDGACAGCAQPPGPPAGRGSGGPA 168
Db |||||
578 GVMGFPFGKGAAGFPKAGRGVPGPPGAVGPAKDGACAGCAQPPGPPAGRGSGGPA 637
Qy 169 GSPGFGILGPPAGPPGAGKRGEGVPGDILGAGPSGPAE-----PGP----- 212
Db |||||
638 GSPGFGILGPPAGPPGAGKRGEGVPGDILGAGPSGARGRFPFGRGVQGPFGAGPR 697
Qy 213 -----TGLPGPPGRRGGSGRFPQADGAVAGPKGPPAGRGSGFPAGPKGSP 258
Db |||||
698 GAGAPGNDGAKGADAGAPGSGQAGLQGMPCERGAAGLPGPKGDRGADGPKGADGSP 757
Qy 259 GE-----AGRPBAGLPGAK-----GLTSGPSGPDGKTGTPPGAGQDGRP 300
Db |||||
758 KGDGVRGLGTGIGPPGAPAGDGGESGSPGAGTGAAGAPGDRGEPGPPGAGPAGPP 817
Qy 301 GPPGPPGARGO-----AGVVGFPKGAAGEPKAGRGVPGPPGA----- 341
Db |||||
818 GADQPAKAGPPGADAGKADGAGPPGAPGAGPPGPIGNVAPGAKGARGAGPPGATGFP 877
Qy 342 -----VGPAGKDGAGAGPPGAPG---AGRGEGQAGSPGQGLPGPAGPPGAGKP 393
Db |||||
878 GAAGRVGPPGPGSNAGPPGPPGACKEGKGRGTCGPRGTCGPRGTCGPRGTCGPRG 937
Qy 394 GEQGVGDLGAPGSGAGRPGLPGPCRGSGRFPGADGAVAGPKGPPAGRGSP 453
Db |||||
938 GADGAPAGPPGPPGQIAGRGVVLPGQGRGEPFLPGPSGPPGQSGASGAGRPP 997
Qy 454 GPAGP---KSPGAGRPAGLPGAKGLTSGPSGPDGKTGTPPGAGQDGRPPGPPGP 510
Db |||||
998 GPMGPPGLAGPPGSGREGAPGASGSPGRDGSPPGAKGDRGTGTPGAGPPGAPGAPGV 1057
Qy 511 GARGQAGVMGPPGPKGAAGEPKAGRGVPGPPGAVGPAKDG-----AGAQ 558
Db |||||
1058 GPAGKSGDRGTGTPAGPAGPVGPPAGARGPAGPQGRGDKGTGQSGDRGKGRGSG 1117
Qy 559 GPPGP-----AGPAGRGEGQAGSP---GFQGLPGPAGPPGAGKRGEGVP 603
Db |||||
1118 GPPGPPSGGQSGASGAPGPPGSPGAGAPGKDGKGLGPGIGPPGPRGTGDAGPV 1177
Qy 604 GDLAGPSSGAPG 616
Db |||||
1178 GPPGPPGPPGPPG 1190

RESULT 2
ID O76045 PRELIMINARY; PRT; 1461 AA.
AC O76045;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Pro alpha 1(I) collagen.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```

```
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
[2]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=88329734; PubMed=2843432;
RA D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT "Complete nucleotide sequence of the region encompassing the first
twenty-five exons of the human pro alpha 1(I) collagen gene.";
RL Gene 67:105-115(1988).
[3]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
Jaanisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
[4]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=91138770; PubMed=1995349;
RA Maatta A., Bornstein P., Penttinen R.P.;
RT "Highly conserved sequences in the 3'-untranslated region of the
COL1A1 gene bind cell-specific nuclear proteins.";
RL FEBS Lett. 279:9-13(1991).
[5]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=92157916; PubMed=1787829;
RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL Matrix 11:375-379(1991).
[6]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=98107942; PubMed=9443882;
RA Korkko J., Ala-Kokko L., De Paape A., Nuytinck L., Earley J.,
Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
scanning by conformation-sensitive gel electrophoresis identifies only
COL1A1 mutations in 15 patients with osteogenesis imperfecta type I;
identification of common sequences of null-allele mutations.";
RL Am. J. Hum. Genet. 62:98-110(1998).
[7]
RN RP SEQUENCE FROM N.A.
RX Korkko J.M., Earley J.J., Nuytinck L., DePaape A., Prockop D.J.,
Ala-Kokko L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017178; AAB94054.2; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0003737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PSS0184; VWF_C_2; 1.
KW Collagen.
SQ SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30EA78E21 CRC64;
64.0%; Score 2233.5; DB 2; Length 1461;
```

Query Match

Best Local Similarity 60.0%; Pred. No. 1.4e-82;  
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGERGSGRFPFGADGVAGPKGAGRGSGPQA 48  
DB 455 GPPGAGBEGKRGARGEPTGLPGERGSGRFPFGADGVAGPKGAGRGSGPQA 514  
QY 49 GPKSGPAGRPGEAGLFGAKGLTGSPGSPGDKTGTTPGAGDGRGPPGPPGARGQA 108  
DB 515 GPKSGPAGRPGEAGLFGAKGLTGSPGSPGDKTGTTPGAGDGRGPPGPPGARGQA 574  
QY 109 GVMGPPGKAGBPKKAGRGVPPGCAVCPAGKDGAGAGNQGPPGAGPAGERGEGQA 168  
DB 575 GVMGPPGKAGBPKKAGRGVPPGCAVCPAGKDGAGAGNQGPPGAGPAGERGEGQA 634  
QY 169 CSPGFGQLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAGE-----PGP--- 212  
DB 635 CSPGFGQLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAGE-----PGP--- 212  
QY 213 -----TGLPGERGSGRFPFGADGVAGPKGAGRGSGPQA 694  
DB 695 GAGAPGNDGAKGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 754  
QY 259 GE-----AGRPAGAGLFGAK-----GLTSGSPGSPGDKTGTTPGAGDGRGPP 300  
DB 755 GKDGVRLGTGTPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 814  
QY 301 GPPGPPGARGQ-----AGVMFPKPKGAAGPAGPAGPAGPAGPAGPAGPAGPAGP 341  
DB 815 GADGPKGAKGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 874  
QY 342 -----VGPAGKDGAGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 393  
DB 875 GAAGRVGPPGSGNAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 934  
QY 394 GEGVPGDLGAPGSPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 453  
DB 935 GADGAPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 994  
QY 454 GPAGP-----KSPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 510  
DB 995 GPMGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1054  
QY 511 GARGGAGVMPGPPGKAGBPKKAGRGVPPGCAVCPAGKDGAGAGNQGPPGAGPAGERGEGQA 558  
DB 1055 GPAGKSGDRGTGTGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1114  
QY 559 GPPGP-----AGPAGRGEGGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 603  
DB 1115 GPPGPPGSPGQPSGASGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1174  
QY 604 GDLGAPGSPGAG 616  
DB 1175 GPPGPPGPPGPPG 1187

RESULT 3  
Q8N473 PRELIMINARY; PRT; 1464 AA.  
AC Q8N473;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Alpha 1 type I collagen, preproprotein.  
GN Name=COL1A1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Prange C.,  
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Souffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RC TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; SC036533; AAH36531.1; -  
DR GO; GO:0005581; C:cytoplasm; IEA.  
DR GO; GO:0005737; C:cytoplasm; IEA.  
DR GO; GO:0005203; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0005817; F:phosphate transport; IEA.  
DR InterPro; IPR008161; Clg helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR000885; Fib collagen\_C.  
DR InterPro; IPR009041; PMF\_SGCI.  
DR InterPro; IPR001007; VWF\_C.  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 18.  
DR ProDom; PD000007; Clg helix; 3.  
DR ProDom; PD002078; Fib collagen\_C; 1.  
DR SMART; SM00038; COLFI; 1.  
DR SMART; SM00214; VWF; 1.  
DR PROSITE; PS01208; VWF\_C; 1.  
DR PROSITE; PS0184; VWF\_C; 1.  
KW Collagen.  
SQ SEQUENCE 1464 AA; 139010 MW; B0581F8D1C89DDE8 CRC64;

Query Match 64.0%; Score 2233.5; DB 2; Length 1464;  
Best Local Similarity 60.0%; Pred. No. 1.4e-82;  
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGERGSGRFPFGADGVAGPKGAGRGSGPQA 48  
DB 458 GPPGAGBEGKRGARGEPTGLPGERGSGRFPFGADGVAGPKGAGRGSGPQA 517  
QY 49 GPKSGPAGRPGEAGLFGAKGLTGSPGSPGDKTGTTPGAGDGRGPPGPPGARGQA 108  
DB 518 GPKSGPAGRPGEAGLFGAKGLTGSPGSPGDKTGTTPGAGDGRGPPGPPGARGQA 577  
QY 109 GVMGPPGKAGBPKKAGRGVPPGCAVCPAGKDGAGAGNQGPPGAGPAGERGEGQA 168  
DB 578 GVMGPPGKAGBPKKAGRGVPPGCAVCPAGKDGAGAGNQGPPGAGPAGERGEGQA 637  
QY 169 CSPGFGQLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAGE-----PGP--- 212  
DB 638 CSPGFGQLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAGE-----PGP--- 212  
QY 213 -----TGLPGERGSGRFPFGADGVAGPKGAGRGSGPQA 697  
DB 698 GAGAPGNDGAKGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 757  
QY 259 GE-----AGRPAGAGLFGAK-----GLTSGSPGSPGDKTGTTPGAGDGRGPP 300  
DB 758 GKDGVRLGTGTPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 817

QY 301 GPPPPGARGQ-----AGVMGFGPKCAAGPGKACGERGVPPGPA----- 341.  
Db 818 GADGQPKAGBPGDAGAKGDAGPPGAPGPGPIGNVGAAPKAGARGAGPPGATGFP 877  
QY 342 -----VGPAGKDGAGAGQGGPPGAPG---AGERGEQPGAGSPGQGLPGPAGPPGAGK 393  
Db 878 GAAGRVGPPGSGNAGPPGPPGAPGKGGKGRGETGAPGPGVEGPPGPPGAGKGGSP 937  
QY 394 GEQVPGDGLGAPSGPAGSPGPTGTLPGPGRGPGSRGFPAGDVAGVAGKGPAGRGSP 453  
Db 938 GADGPAGAPGTPGQGIAGQGVVGLPGQGRGPGFPLPGPSGPGKQSGSGASGERGPP 997  
QY 454 GPAGP---KSGPGAGRPGEAGLPGAKGLTGSPGPGDGTGTPPGPAGQDGRPPGPP 510  
Db 998 GPMGPPGLAGPPGSGREGAPCAEGSPGRDGSPOAKGDRGETGAPGPGAPGAPGVP 1057  
QY 511 GARGQAGVMGPPGKGAAGPGKAGRGVPGPPGAVGAPGKDBE-----AGAQ 558  
Db 1058 GPAGKSGDRGTGAPGAPGVPVPGARGPAGPQGRGDKGTGEBQGRGKIGHRGFSLG 1117  
QY 559 GPPGCP-----AGPAGERGEGPAGSP---CFQGLPGPAGPPGAGKGGVGP 603  
Db 1118 GPPGPPGSGPQSGSGAGPAGPRGPGSAGAPKQGLNGLPDPIGPPGPRGTGDRGVP 1177  
QY 604 GDLGAPGSPGAP 616  
Db 1178 GPPGPPGPPGPPG 1190

RESULT 4  
CALL CANFA  
ID CALL CANFA STANDARD; PRT; 1460 AA.  
AC Q9XJ7;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Collagen alpha 1(I) chain precursor.  
GN Name=COL1A1;  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=96115;  
RN [1]  
RP SEQUENCE FROM N.A., AND VARIANT OF ALA-208.  
RC TISSUE=Skin;  
RX MEDLINE=21023337; PubMed=11147834;  
RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;  
RT "Sequence of normal canine COL1A1 cDNA and identification of a  
RT heterozygous alpha1(I) collagen Gly208Ala mutation in a severe case of  
RT canine osteogenesis imperfecta";  
RL Arch. Biochem. Biophys. 384:37-46 (2000).  
CC -!- FUNCTION: Type I collagen is a member of group I collagen  
CC (fibrillar forming collagen).  
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta  
CC (OI).  
CC -!- SIMILARITY: Belongs to the fibrillar collagen family.  
CC -!- SIMILARITY: Contains 1 VWF domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC  
CC EMBL; AF153062; AAD34619.1; .  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.

DR InterPro; IPR00885; Fib collagen\_C.  
DR InterPro; IPR001007; VWF\_C.  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 18.  
DR ProDom; PD000007; C1g\_helix; 2.  
DR ProDom; PD002078; Fib collagen\_C; 1.  
DR PROSITE; PS01208; VWF\_C; 1.  
DR PROSITE; PS0184; VWF\_C; 2; 1.  
KW Collagen; Connective tissue; Disease mutation; Extracellular matrix;  
KW Glycoprotein; Hydroxylation; Pyrrolidone carboxylic acid; Repeat;  
KW Signal.  
FT SIGNAL 1 22 By similarity.  
FT PROPEP 23 157 N-terminal propeptide.  
FT CHAIN 158 1214 Collagen alpha 1(I) chain.  
FT PROPEP 1215 1460 C-terminal propeptide.  
FT DOMAIN 34 92 VWF.  
FT DOMAIN 158 174 Nonhelical region (N-terminal).  
FT DOMAIN 175 1188 tripe-helical region.  
FT DOMAIN 1189 1214 Nonhelical region (C-terminal).  
FT SITE 741 743 Cell attachment site (Potential).  
FT SITE 1089 1091 Cell attachment site (Potential).  
FT MOD\_RES 158 158 Pyrrolidone carboxylic acid (By similarity).  
FT MOD\_RES 166 166 Allylsine (By similarity).  
FT MOD\_RES 261 261 5-hydroxylysine (By similarity).  
FT CARBOHYD 1160 1160 3-hydroxyproline (By similarity).  
FT CARBOHYD 261 261 O-linked (Gal.. ) (By similarity).  
FT CARBOHYD 1361 1361 N-linked (GlcNAc.. ) (By similarity).  
FT VARIANT 208 208 G -> A (in OI; severe).  
SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;

Query Match 64.0%; Score 2231.5; DB 1; Length 1460;  
Best Local Similarity 60.2%; Pred. No. 1.7e-82;  
Matches 441; Conservative 20; Mismatches 155; Indels 117; Gaps 12;

QY 1 GPP-----GEPGTGLPGPPQGERGPGSRGPGFADGAGVAGPKPAGERGSPGPA 48  
Db 454 GPPGAGEGKRGAGEPGTGLPPPERGPGRGPGFADGAGVAGPKPAGERGSPGPA 513  
QY 49 GPKSGPGEAGRPCEAGLPGAKGLTSPSGPGDKTGPFPAGQDGRPGPPGARGQA 108  
Db 514 GPKSGPGEAGRPCEAGLPGAKGLTSPSGPGDKTGPFPAGQDGRPGPPGARGQA 573  
QY 109 GVMGPPGKGAAGEPKAGERGVPPGAVGAGVAGDGEAGAGOGPPGAPGAGERGQGA 168  
Db 574 GVMGPPGKGAAGEPKAGERGVPPGAVGAGVAGDGEAGAGOGPPGAPGAGERGQGA 633  
QY 169 GSPGFQGLPGPAGPPGAEAGKPGEQVPGDLGAPSGPAGE-----PGP---- 212  
Db 634 GSPGFQGLPGPAGPPGAEAGKPGEQVPGDLGAPSGPAGEGPGERGVQGPAGPR 693  
QY 213 -----TGLPDPGERGPGSRGPGFADGAGVAGPKPAGERGSPGPKGSP 258  
Db 694 GANGAPGNDGAKGDAGAPAGSQAPLQGMFGERGAAGLPGPKGDRDAGPKGADGSP 753  
QY 259 GE-----AGRPGEAGLPGAK-----GLTSGPSGPGDKTGPFPAGQDGRP 300  
Db 754 GKDVRLTGTGTPGPPGAPAGKXGAGSPGAGTGARGAGDGRGPGPPGAPGAPFP 813  
QY 301 GPPPPGARGQ-----AGVMGFGPKCAAGPGKACGERGVPPGPA----- 341  
Db 814 GADGQPKAGBPGDAGAKGDAGPPGAPGPGPIGNVGAAPKAGARGAGPPGATGFP 873  
QY 342 -----VGPAGKDGAGAGQGGPPGAPG---AGERGEQPGAGSPGQGLPGPAGPPGAGK 393  
Db 874 GAAGRVGPPGSGNAGPPGPPGAPGKGGKGRGETGAPGPGVEGPPGPPGAGKGGSP 933  
QY 394 GEQVPGDGLGAPSGPAGSPGPTGTLPGPGRGPGSRGFPAGDVAGVAGKGPAGRGSP 453  
Db 938 GADGPAGAPGTPGQGIAGQGVVGLPGQGRGPGFPLPGPSGPGKQSGSGASGERGPP 993  
QY 454 GPAGP---KSGPGAGRPGEAGLPGAKGLTGSPGPGDGTGTPPGPAGQDGRPPGPP 510

```

Db 994 CFMGPGLAGPPGESRGSPGCAESGCRDGSPPGKGDRTGTGAGPPGAPGAPGPV 1053
QY 511 GARGGAGVMPGPKGAAGEFGKAGRGVPGPPGAVGAGPKDGE-----AGAG 558
Db 1054 GPAGKNGDRGETGAPGAPGTPGVGARGPAGPQGRGDKGTGTGQGRGKXGHRGFSGLQ 1113
QY 559 GPPGP-----AGPAGERGEQGPAGSP---GFQGLPGPAGPPGAGKPGEGQGV 603
Db 1114 GPPGPPGSPGQGGSPGAGPAGPPGAGSPGKGLNGLPGLGPPGPRGRTGDAGPV 1173
QY 604 GDLGAPGSPGAP 616
Db 1174 GPPGPPGPPGPPG 1186

RESULT 5
Q6LAN8
ID Q6LAN8 PRELIMINARY; PRT; 1069 AA.
AC Q6LAN8;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DE Collagen type I alpha 1 (Fragment).
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97141927; PubMed=8988177;
RA Simon M., Pedetour F., O'Brien K., Sirvent N., Kedra D., Guilbaud C.,
RA Turc-Carel C., Dumanski J.P.;
RT "Regulation of the platelet-derived growth factor B-chain gene via
RT fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans
RT and giant-cell fibroblastoma.";
RL Nat. Genet. 15:95-98(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA O'Brien K.P.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X98705; CAA67261.1; -.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR009041; PWE_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01391; Collagen; 15.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PDO00007; C1g_helix; 3.
DR SMART; SMO0214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS50184; VWF_C_2; 1.
KW Collagen.
FT NON_TER 1069 1069
SQ SEQUENCE 1069 AA; 97445 MW; BE279B10572FB980 CRC64;

Query Match 63.7%; Score 2222; DB 2; Length 1069;
Best Local Similarity 57.6%; Pred. No. 3.3e-82;
Matches 448; Conservative 17; Mismatches 151; Indels 162; Gaps 15;

QY 1 GPPGEPGTGLPGPPGRRGGSGRFFGADGVGPKGAPAGERSPPGAPGKGSFGGA--- 57
Db 179 GPMGSPGRLGPPGAPGPPGQGGPGEFGAGSGFMGPRGPPGPKNGKDDGAGKP 238
QY 58 GRPGEAGLP---GAKGLTSGSPGSP-----DGKTPGPAGQDGRPPGPPGAR 105
Db 239 GRPGERPPGQARGLPGTAGLFGMKHGFSLDCAKGDAGPAGPKGPFSGFNGAP 298
QY 106 GQAGVMGPPGPKGAAGBFGKAGER-----GVPGPPGAVGPA-----GKDGAGAG 150
Db 299 GQMGFRLGPRGRFPGAPGAGAGNDGATCAAGPPGTGTPAGPPGPPGAVGAKGAGPQ 358
QY 151 GPPGAPGAGRGQGP-----AGSPGFGQLGPA 180

```

```

Db 359 GPRGSEGGVGRGEPGPPGAPAGNPGADGQPGAKGANGAPGAGPFGARGPS 418
QY 181 GPPGAGRKGEGVPGDILGAP-----GPSGRA-----GERGPT 213
Db 419 GPGGPGPPGPKNGSGEFGAPGSKGDTAKGEPGPGVQGGPPGAGEGKGAERGEP 478
QY 214 GLPGPPGRRGGSGRFFGADGVGPKGAPAGERSPPGAPGKGSFGGAGRGAGLPGAK 273
Db 479 GLPGPPGRRGGSGRFFGADGVGPKGAPAGERSPPGAPGKGSFGGAGRGAGLPGAK 538
QY 274 GLTSGPSPGPDGKTPGPPGAGQDGRPPGPPGARGAGVGMGPPGKGAAGFPGKAGER 333
Db 539 GLTSGPSPGPDGKTPGPPGAGQDGRPPGPPGARGAGVGMGPPGKGAAGFPGKAGER 598
QY 334 GVPGPPGAVGPAGKDEAGAGQPPGAPGAGERGEGQPGAGSPGFOGLPGPAGPGEAGKP 393
Db 599 GVPGPPGAVGPAGKDEAGAGQPPGAPGAGERGEGQPGAGSPGFOGLPGPAGPGEAGKP 658
QY 394 GEQVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
Db 659 GEQVPGDLGAPGSPGAGERGEGFPGERGVPQPPGPPGAPRGANGAPNDGAKGDAGAPGAP 718
QY 424 GERGGSGSGPPGADGVGPKGAPAGERSPPGAPGKGSFG-----AGRGAGLP 474
Db 719 GSGGAPGLQGMPPGERGAAGLPQPKGDRDAGPKGADGSGDKGVRLTGTPIGPPGAPAG 778
QY 475 GAK-----GLTSGSPSPDGKTPGPPGAGQDGRPPGPPGARGQ----- 515
Db 779 GDKGSGSPGAPGTCARGAPDGRGEPGPPGAPGADGQPGAKGEGPDAGAKGDA 838
QY 516 -----AGVMGPPGKGAAGRGKAGERGVPPGPA-----VGPAGKDEAGAGP 561
Db 839 GPPGAPGAPGPPGPIGNVGAFAKAGSAGSPGATGFPGAAGRVGPPGSGNAGPPGPP 898
QY 562 GPAGS---AGERGEGQPGAGSPGFOGLPGPAGPPEAGKPGQGVGDLGAPGSPGAPG 616
Db 899 GPAGKEGKGRGTGTPAGRGVEGVPGPPGAPGKGSFGADGPPAGPPTGPGQGIAG 956

RESULT 6
CAA67261
ID CAA67261 PRELIMINARY; PRT; 1069 AA.
AC CAA67261;
DT 02-MAR-2004 (TRENBLrel. 27, Created)
DT 02-MAR-2004 (TRENBLrel. 27, Last sequence update)
DE Collagen type I alpha 1 (Fragment).
GN COL1A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Simon M., Pedetour F., O'Brien K., Sirvent N., Kedra D., Guilbaud C.,
RA Turc-Carel C., Dumanski J.P.;
RT "Regulation of the platelet-derived growth factor B-chain gene via
RT fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans
RT and giant-cell fibroblastoma.";
RL Nat. Genet. 15:95-98(1997).
DR EMBL; X98705; CAA67261.1; -.
FT NON_TER 1069 1069
SQ SEQUENCE 1069 AA; 97446 MW; BE279B10572FB980 CRC64;

Query Match 63.7%; Score 2222; DB 2; Length 1069;
Best Local Similarity 57.6%; Pred. No. 3.3e-82;
Matches 448; Conservative 17; Mismatches 151; Indels 162; Gaps 15;

QY 1 GPPGEPGTGLPGPPGRRGGSGRFFGADGVGPKGAPAGERSPPGAPGKGSFGGA--- 57
Db 179 GPMGSPGRLGPPGAPGPPGQGGPGEFGAGSGFMGPRGPPGPKNGKDDGAGKP 238
QY 58 GRPGEAGLP---GAKGLTSGSPGSP-----DGKTPGPAGQDGRPPGPPGAR 105
Db 239 GRPGERPPGQARGLPGTAGLFGMKHGFSLDCAKGDAGPAGPKGPFSGFNGAP 298
QY 106 GQAGVMGPPGPKGAAGBFGKAGER-----GVPGPPGAVGPA-----GKDGAGAG 150
Db 299 GQMGFRLGPRGRFPGAPGAGAGNDGATCAAGPPGTGTPAGPPGPPGAVGAKGAGPQ 358
QY 151 GPPGAPGAGRGQGP-----AGSPGFGQLGPA 180

```







Db 687 GNNGAPGNDGAGTGTGAPGAGSCAPGQLQMPGGERGAAGLPGPKGDRGADGPKGADGSP 746  
Qy 259 GE-----AGRPGAGLPGAGLGTGSPSPGPD-----GKTGPPGAGQDGRP 300  
Db 747 KDGAGLGTGIPGPPGAGAGEDGAGSPGPPGTGARGAPGDRGAGPPGAGFAGPP 806  
Qy 301 GPPGPPGARGO-----AGVMGFPKGAAGPPGKAGRGVPGPPGA----- 341  
Db 807 GADGQPAKGFBDGTGVKGDAGPPGAPGPPGPIGNVAGPAPKGRGAAGPPGATGFP 866  
Qy 342 -----VGPAGKDGAGAGCGPPGAPG---AGERGQAGPAGSPGQGLPGPAGPPGAGK 393  
Db 867 GAAGRVGPPGSPGAGPPGPPGPKGKGKGRGTGAPGPPGPPGPPGPPGAGK 926  
Qy 394 CEQGVGDLGAPGSPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 453  
Db 927 GADGAPGSPGPPGQIAGRGVGLPGQRCGERGPPGLPGSPGPPGPPGPPGPPG 966  
Qy 454 GPAGP---KGSFGAGRGAGLPGCAKLTGSPSPGPDGKTGPPGAGQDGRGPPGPP 510  
Db 987 GPMGPPGLAGPPGSPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1046  
Qy 511 GARGQAGVMGPPGPKGAGGPPKAGRGVPPG-----PGAV 546  
Db 1047 GPAGXNDRGTGTGAPGAPGPPGAPGAPGPPGPPGPPGPPGPPGPPGPPG 1106  
Qy 547 GPAGKDXGAGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 603  
Db 1107 GPPSPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1166  
Qy 604 GDLGAPGPPGPPG 616  
Db 1167 GPPGPPGPPGPPG 1179

RESULT 9

Q810J9 PRELIMINARY; PRT: 1453 AA.  
AC Q810J9;  
DT 01-JUN-2003 (TRENBLrel. 24, Created)  
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE Procollagen, type I, alpha 1.  
GN Name=Colla1;  
OS Mus musculus (Mouse)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N; TISSUE=Colon;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RN SEQUENCE FROM N.A.  
RP STRAIN=FVB/N; TISSUE=Colon;  
RC Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050014; AAH50014.1; -  
DR MGD; MGI:88467; Coll1a1.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005737; C:cyclopasim; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0006817; P:phosphate transport; IEA.  
DR InterPro; IPR008161; Clg helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR000885; Fib collagen\_C.  
DR InterPro; IPR009041; FMP\_SGCI.  
DR InterPro; IPR001007; VWF\_C.  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 18.  
DR Pfam; PF00093; VWC; 1.  
DR ProDom; PD000007; Clg helix; 3.  
DR ProDom; PD002078; Fib collagen\_C; 1.  
DR SMART; SM00038; COLFI; 1.  
DR SMART; SM00214; VWC; 1.  
DR PROSITE; PS01208; VWF\_C\_1; 1.  
DR PROSITE; PS01184; VWF\_C\_2; 1.  
KW Collagen.  
SQ SEQUENCE 1453 AA; 138032 MW; 0B7F06BBB9A1D5EA CRC64;  
  
Qy 1 GPP-----GEPGPTGLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 48  
Db 447 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 506  
Qy 49 GPKGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 108  
Db 507 GPKGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 566  
Qy 109 GVMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 168  
Db 567 GVMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 626  
Qy 169 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 212  
Db 627 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 686  
Qy 213 -----TGLPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 258  
Db 687 GNNGAPGNDGAGTGTGAPGAGSCAPGQLQMPGGERGAAGLPGPKGDRGADG 746  
Qy 259 GE-----AGRPGAGLPGAGLGTGSPSPGPD-----GKTGPPGAGQDGRP 300  
Db 747 KDGAGLGTGIPGPPGAGAGEDGAGSPGPPGTGARGAPGDRGAGPPGAGFAGPP 806  
Qy 301 GPPGPPGARGO-----AGVMGFPKGAAGPPGKAGRGVPGPPGA----- 341  
Db 807 GADGQPAKGFBDGTGVKGDAGPPGAPGPPGPIGNVAGPAPKGRGAAGPPGATGFP 866  
Qy 342 -----VGPAGKDGAGAGCGPPGAPG---AGERGQAGPAGSPGQGLPGPAGPPGAGK 393  
Db 867 GAAGRVGPPGSPGAGPPGPPGPKGKGKGRGTGAPGPPGPPGPPGPPGAGK 926  
Qy 394 CEQGVGDLGAPGSPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 453  
Db 927 GADGAPGSPGPPGQIAGRGVGLPGQRCGERGPPGLPGSPGPPGPPGPPGPPG 966  
Qy 454 GPAGP---KGSPPGAGRGAGLPGCAKLTGSPSPGPDGKTGPPGAGQDGRGPPGPP 510  
Db 987 GPMGPPGLAGPPGSPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1046  
Qy 511 GARGQAGVMGPPGPKGAGGPPKAGRGVPPG-----PGAV 546



```
RC STRAIN=C57BL/6J; TISSUE=Mammary tumor;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klatte R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Haieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stappleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Paney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary tumor;
RA Strausberg R.;
RL Submitted (Oct-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC059281; AAH59281.1; --
SQ SEQUENCE 1225 AA; 117860 MW; B6B86CB4457F4D9 CRC64;

Query Match 62.4%; Score 2176.5; DB 2; Length 1225.
Best Local Similarity 56.7%; Pred. No. 2.4e-80;
Matches 431; Conservative 28; Mismatches 157; Indels 144; Gaps 13;

QY 1 GPPGPGTGLPDPGE-----RCGPGSRGPGADGVAGPKGPRGERSGPA 48
DB 180 GPPGAPGQGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPG 239
QY 49 GPKGPGGAGR-----CEAGLPKAGLGTSPGSPGPKGTGPPGPAQD 93
DB 240 GARGLPGTAGLPCKMKGHRGSLDCAKGDAGPAGPKGPGSPGPGAGQMGPRGLPGER 299
QY 94 GRPGPPGPPGARGQAGVGMGPPGPKAAGE-----PGKAGRGVPPGPGVAGVPKAGKGEA 147
DB 300 GRPGPPGTAGRGNDGAVGAGPPGPGTGTGTPGPGFPGVAGKAGAGPGQARGSGPGQVR 359
QY 148 GAQGGPPGAPGAGEGEP-----AGSPGFGQLPGPAPGPGGPKGQEQ 192
DB 360 GEPGPPGAGAGPAGNPGADGQPKAKGANGAPGAGPFGKAGRSGPGSPGPGPK 419
QY 193 GVPGLGAP-----GPSGPA-----GPGPTGLPDPGGERGP 225
DB 420 GNSGEPGAPGKGTGAKGPGATGVQGPFGPGAGEGKRGARGEPGSGLPDPGGERGP 479
QY 226 GSRGPPGADGVAGPKGAGERSGPGAPKSGPAGPGEAGLPKAGLGTSPGSPGPD 285
DB 480 GSRGPPGADGVAGPKGPGSAGPAGPAGPKGSPGAGRPAGLPGKGLGTSPGSPGPD 539
QY 286 GKTGPPGAGQDGRGPPGPPGARGQAGVGMGPPGPKAAGEBFGKAGRGVPPGPAVGA 345
DB 540 GKTGPPGAGQDGRGPPGPPGARGQAGVGMGPPGPKGTAGBFGKAGRGVPPGPAVGA 599
QY 346 GKDGAGAGGPPGAPGAGEGQGPAGSPGPGQGLPGAPGPGGKAGKGEQGVPPDLGAP 405
DB 600 GKDGAGAGGAPGAPGAGEGQGPAGSPGPGQGLPGAPGPGGKAGKGEQGVPPDLGAP 659
QY 406 GPSGPGAGE-----PGP-----TGLPDPGGERGPGSRGFP 435
DB 660 GPSGARGERGPPGRCVQGPFGPAGPRGNAGPNDGAKGDTGAGAGSQAGPGLQMP 719
QY 436 GADGVAGPKGAPGERSGPGAPKSGPGE-----AGRPGGAGLPGKAGLGTSPGSP 486
DB 720 GERGAAGLPKGRDAGPKGADGSPGDKGARGLGTGTPGPPGAPGADKGEAGSPGP 779
```

```
QY 487 GPDGKTGPPGAGQDGRGPPGPPGARGQAGVGMGPPGKAGRGKAGRGVPPGPAV 546
DB 780 GFTGARGAPGDRGEAGPPGPPGAPGAPGAPGPPGAPGAPGAPGAPGAPGAPGAP 839
QY 547 GPAGKDGAGAGQPPGPPGAPGAGE-----QGPAGSPFGQGLPQPA----- 588
DB 840 GPAGARGPAGPGPRGDKGTGEGQDGRGKGRGFGSLGQPPGSPGSPGQSPGASGPA 899
QY 589 -----GPPGAGKGE-----QVPPGDLGAPG-----SGPAG 616
DB 900 GPRGPPGASGSPGDKLNLGPGPIGPPGPRGRTGDSGPAG 939
```

## RESULT 12

```
CALL_CHICK STANDARD; PRT; 1453 AA.
ID P02457;
AC 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Names=COL1A1;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 1-153 FROM N.A.
RX MEDLINE=88056316; PubMed=3678934;
RA Finer M.H., Boedtker H., Doty P.;
RT "Construction and characterization of cDNA clones encoding the 5' end
RT of the chicken pro alpha 1(I) collagen mRNA.";
RL Gene 56:71-78(1987).
RN [2]
RP SEQUENCE OF 1-144 FROM N.A.
RX MEDLINE=88007542; PubMed=2820966;
RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RT "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [3]
RP SEQUENCE OF 152-1187.
RX MEDLINE=82231995; PubMed=7093229;
RA Higberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M., Kang A.H.,
RA Gröss J.;
RT "Amino acid sequence of chick skin collagen alpha 1(I)-C88 and the
RT complete primary structure of the helical portion of the chick skin
RT collagen alpha 1(I) chain.";
RL Biochemistry 21:2048-2055(1982).
RN [4]
RP SEQUENCE OF 1200-1205.
RX MEDLINE=72243016; PubMed=5047697;
RA Eyre D.R., Glincher M.J.;
RT "Evidence for a previously undetected sequence at the carboxyterminus
RT of the alpha 1 chain of chicken bone collagen.";
RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN [5]
RP SEQUENCE OF 981-1453 FROM N.A.
RX MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
RT alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RT including the carboxy-terminal propeptide sequences.";
RL Biochemistry 20:1996-1006(1981).
RN [6]
RP SEQUENCE OF 1311-1453 FROM N.A.
RX MEDLINE=80134546; PubMed=6987088;
RA Shewalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
RA Pastan I., Decrombrughe B., Fietzek P.P., Olsen B.R.;
RT "Nucleotide sequence of a collagen cDNA-fragment coding for the
RT carboxyl end of pro alpha 1(I)-chains.";
```



DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR00885; Fib collagen\_C.  
DR InterPro; IPR001007; VWF\_C.  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 18.  
DR ProDom; PD000007; Clg\_helix; 4.  
DR ProDom; PD002078; Fib\_collagen\_C; 1.  
DR SMART; SM00038; COLFI; 1.  
DR SMART; SM00214; VWC; 1.  
DR PROSITE; PS01208; VWF\_C; 1; UNKNOWN\_1.  
DR PROSITE; PS50184; VWF\_C; 2; 1.  
DR Collagen.  
KW COLLAGEN.  
SQ SEQUENCE 1450 AA; 137563 MW; ABF8A74841B87B7C CRC64;  
  
Query Match 58.6%; Score 2043.5; DB 2; Length 1450;  
Best Local Similarity 54.7%; Pred. No. 5.7e-75;  
Matches 401; Conservative 31; Mismatches 184; Indels 117; Gaps 11;  
  
QY 1 GPP-----GEFGTGLPGPPGGERGSGRGGPGGADGVAGPKGAGERGSGPPA 48  
DB 444 GPPGSGEGRKGSGRGGEPGAPGPPGAGERGSGRGGPGSDGASGPKGAPGERSGVGPA 503  
QY 49 GPKGSPGAGRPGRAGLPCAKGLTSGSGSPGDKTGTGPPAGQDGRPPGPPGARGQA 108  
DB 504 GPKGSGRSGRPGRPGCLPCKAKGLTSGSGSPGDKTGTGAGAGQDGHGPPGSGARGQS 563  
QY 109 GVMGPPGKGAAGEFGKAGERGVPPGAVGPGAGKDGAGACGPPGPPGAGERGEGQGA 168  
DB 564 GVMGPPGKGAAGEFGKAGERGVPPGAVGPGATGAPKDGAGACGPPGPPGSGRGERGEGQGA 623  
QY 169 GSPGQGLPGPAGPGEAGKGEQGVGDGLGAPGSGPAGE-----PGPTGLPGP- 218  
DB 624 GSPGQGLPGPAGPGEAGKGEQGVGDGLGAPGSGPAGE-----PGPTGLPGP- 218  
QY 219 -----PGERGGSGRGGP----- 231  
DB 684 GSPGSGPNDGAKGEAGAGACGGRGPPGLQGMPGERSGAGMPGAKGDRGDATGKADGAP 743  
QY 232 ----GADGVAGPKGAGERGSP-----GPAGKSGPGEAGRGEAGLPCAKGLTSP 279  
DB 744 GKDAGRLGTGTPGPPGAPGDKGEGGSGPAGTARGSGFERGEPGAPGAGICGPP 803  
QY 280 GS----PGPDGKTGPPGAGQDGRPPGPPGARGQAVMGPPGPKGAAGEFGKAGERGV 336  
DB 804 GADGQPKAGEGSDAGPKDAGAPGAPGTGAPGAGNVGAPGPKGTRGAAGPPGATGFP 863  
QY 337 GPPGAVGPKDGEAGAAQGGPPGAPGAGERGQGGAGSGPFGGLPGPAGPP---GEAGKP 393  
DB 864 GAAAGRLGPPGSGNAGPPGPPGPKGKAGKSGRGTGTGAGRSGEPPGAPGPPSGEKSP 923  
QY 394 GEGQVPGDLGAPGSPGAPGPTGLPGPPGRRGGSGRGGPGAGDVAGPKGAGERGSP 453  
DB 924 GSDGPAGAPGIPGPPQIAGQRRGVVGLPGQRRGERSGLPGPAGEPKQGGPNGERGPP 983  
QY 454 GPAGP-----KSGPGEAGRPGAGLPKAG-----GLTSGSPSGPDGKTGPP 495  
DB 984 GFSGPPGLGGPPGPPGREGSPGSEGAFFRDGSPGPKGDRGENGSPGPPGAPGAPGAPV 1043  
QY 496 GPAGQD---GRCPGPPPPGARGQAVMGPPGPKGAAGEFGKAGE-----RGVPPG 543  
DB 1044 GPAGKNGDRGTGTGAPGAPGAPGSGVRGAPGAPGARGDRGAGEGGERGKMGHGRFNGMQ 1103  
QY 544 GAVGPAKDGKGEAGAAQGGPPGAPGAGERGQGGAGSGPFGQLPGPAGPPGEGAGKGEQGV 603  
DB 1104 GPPGPPGSGGEGQAGPSPGAPGPPGSGSGTGKDGVNGLFGIPGPPGRNGDVGPA 1163  
QY 604 GLGAPGSPGAPG 616  
DB 1164 GPDGPPGPPGPPG 1176

RESULT 14  
O93251

O93251 PRELIMINARY; PRT; 1445 AA.  
AC 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Alpha 1 type I collagen.  
GN Name=alpha 1 type I collagen;  
OS Rana catesbeiana (Bull frog);  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae; Rana.  
OX NCBI\_TaxID=8400;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99294154; PubMed=10367734;  
RA Asahina K., Utoh R., Obara M., Yoshizato K.;  
RT "Cell-type specific and thyroid hormone-dependent expression of genes  
of alpha 1(I) and alpha 2(I) collagen in intestine during  
amphibian metamorphosis";  
RL Matrix Biol. 18:89-103(1999).  
DR EMBL; AB015440; BAA28028.1; -;  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005737; C:cytoplasm; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:000617; P:phosphate transport; IEA.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR00885; Fib\_collagen\_C.  
DR InterPro; IPR001007; VWF\_C.  
DR Pfam; PF01410; COLFI; 1.  
DR Pfam; PF01391; Collagen; 18.  
DR ProDom; PD000007; Clg\_helix; 4.  
DR ProDom; PD002078; Fib\_collagen\_C; 1.  
DR SMART; SM00038; COLFI; 1.  
DR SMART; SM00214; VWC; 1.  
DR PROSITE; PS01208; VWF\_C; 1.  
DR PROSITE; PS50184; VWF\_C; 2; 1.  
KW Collagen.  
SQ SEQUENCE 1445 AA; 137251 MW; P59B550C9873F04 CRC64;  
  
Query Match 58.6%; Score 2042.5; DB 2; Length 1445;  
Best Local Similarity 49.7%; Pred. No. 6.2e-75;  
Matches 421; Conservative 34; Mismatches 161; Indels 231; Gaps 18;  
  
QY 1 GPPGEPGPTGLPPGPPGSGSGS-----RGFFGADGVAGPKGPA-----GER 42  
DB 221 GKPGRPGRPPGPPGQARGLPGTAGLPKMGHGRFNGLDGAKGDTGPAGPKGPPGNGEN 280  
QY 43 GSPGAPGPKSGPGEAGRPGPAGLPGAKGLTSGSGSPGDKTGPPE-----A 90  
DB 281 GAFQVGRGLPGERGPPGSGPAGARGNDGTGAAAGPPGPTGTPGPPGPGVGPAGDA 340  
QY 91 GQGRPPPPPPGARGQAVMGPPGPKGAAGEFGKAGERGVPPGAVGPAKDGEGAGQ 150  
DB 341 GPQSRGPDGPPQGNRGEPPGAPGQAGPAGSAGNFGTDGQPKAKGATGAPGAPGPPCAR 400  
QY 151 GPPGAPGAGERGEGQGAPSGPFGQLPGPAGPPGEGAGKPGCEQGVGDGLGAPGSGPAGEP 210  
DB 401 GAPGQPGSGSPGPKGNNGEPGAGKNGEPGAGSGPAGSQGPPGPPGEGKSGRGE 460  
QY 211 GPTGLPGERGPPGSGRGGPPGADGVAGPKGAGERGSGPAGPKSGPGEAGRFGAAGLP 270  
DB 461 GPSPPGAGERGAPGSRGPPGADGACGPKGPPGERGPPVGSAGPKSGSGRPPGEGPL 520  
QY 271 GAKGLTSGSGSPGDKTGTGPPGAGQDGRPPGPPGARGQAVMGPPGPKGAAGEFGKA 330  
DB 521 GAKGLTSGSGSPGDKTGTGPPGAGQDGRPPGPPGARGQAVMGPPGPKGAAGEFGKP 580  
QY 331 GERGVPPGAVGPAKDGAGAGQGGPPGAPGAGERGEGQGAPGSGPFGQLPGPAGPPGEA 390  
DB 581 GERGVPPGAVGPAKDGAGAGQGGPPGAPGAGERGEGQGAPGSGPFGQLPGPAGPPGEA 640  
QY 391 GKPGEQGVGD-----LGAPGSGP-----AGEP 414





*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:22:25 ; Search time 129.602 Seconds  
(without alignments)  
2272.479 Million cell updates/sec

Title: US-10-658-989a-3  
Perfect score: 4640  
Sequence: 1 GPGCEPPTLPGPPGRRG.....GEQGVFDLGAQPGSQPAGG 821

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_238ep04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

ALIGNMENTS

RESULT 1	
ADM48392	
ID	ADM48392 standard; protein; 821 AA.
XX	XX
AC	ADM48392;
XX	XX
DT	03-JUN-2004 (first entry)
XX	XX
DE	Human recombinant gelatin-like polypeptide Hu-4.
XX	XX
KW	Plasma substitute; Gelatin-like protein; plasma expander; human.
XX	XX
OS	Homo sapiens.
XX	XX
EN	EP1398324-A1.
XX	XX
PD	17-MAR-2004.
XX	XX
PF	11-SEP-2002; 2002EP-00078745.
XX	XX
PR	11-SEP-2002; 2002EP-00078745.
XX	XX
PA	(FUJF ) FUJI PHOTO FILM BV.
XX	XX
PI	Bouwstra JB, Toda Y;
XX	XX
DR	WPI; 2004-229415/22.
XX	XX
PT	Composition useful as substitute for plasma, comprises solution of saline
XX	XX
PS	and recombinant gelatin-like protein having colloid osmotic function.
XX	XX
CC	Example 1; SEQ ID NO 3; 31pp; English.

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	DB ID	Description
1	4640	100.0	821	8	ADM48392 Human rec
2	3482	75.0	617	8	ADM48391 Human rec
3	3482	75.0	617	8	ADQ26217 Human gel
4	2765.5	59.6	1449	4	AAE02535 Porcine a
5	2765	59.6	1057	3	AAy84544 A human c
6	2765	59.6	1058	3	AAy84403 Amino aci
7	2765	59.6	1107	3	AAy84540 Amino aci
8	2765	59.6	1171	3	AAy84538 A chimeri
9	2765	59.6	1464	2	AAw68485 Human rec
10	2765	59.6	1464	7	ADD45059 Human pro
11	2765	59.6	1464	7	ADD45055 Human pro
12	2765	59.6	1464	7	ADD45051 Human pro
13	2761	59.5	1461	5	ABG93947 Human pol
14	2761	59.5	1464	4	AAU14136 Human nov
15	2759	59.5	1057	3	AAy84541 Amino aci
16	2757	59.4	1388	3	AAy84539 Amino aci
17	2755	59.4	1464	4	AA82454 Human pro
18	2755	59.4	1464	5	AB590764 Human tum
19	2755	59.4	1464	5	ABp68610 Human pan
20	2755	59.4	1464	6	ABU54471 Human tum
21	2755	59.4	1464	6	ABr47417 Breast ca
22	2755	59.4	1464	6	ABr92064 Human cer
23	2755	59.4	1464	7	ADD14142 Human src
24	2755	59.4	1464	7	ADp65246 Human alp
25	2755	59.4	1464	8	ADQ19470 Human sof

PF	11-SEP-2002; 2002EP-00078745.
XX	
PR	11-SEP-2002; 2002EP-00078745.
XX	
PA	(FUJF ) FUJI PHOTO FILM BV.
XX	
PI	Bouwstra JB, Toda Y;
XX	
DR	WPI; 2004-229415/22.
XX	
PT	Composition useful as substitute for plasma, comprises solution of saline and recombinant gelatin-like protein having colloid osmotic function.
PT	
XX	Example 1; SEQ ID NO 3; 31pp; English.
PS	
XX	The present sequence is the protein sequence of recombinant gelatin-like protein Hu-3. This is a tetramer of human recombinant gelatin-like protein Hu-1 ADM48390. A claimed composition suitable as a substitute for plasma comprises a solution of saline and a protein having a colloid osmotic function. The protein is a recombinant gelatin-like protein with a molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric point of less than 8. It is especially Hu-1 or Hu-deam tetramer of the protein. The recombinant gelatin-like protein, or a dimer, trimer or tetramer of the protein, is useful as a plasma expander that has a lower clearance rate from blood circulation, provides better and predictable regulation of clearance rate and which is less susceptible to proteolytic degradation than presently used gelatin derivatives. Recombinant gelatin-like proteins that are in essence free of hydroxyproline do not give rise to an immunological reaction with blood samples containing IGE antibodies. The gelatin-like proteins can be covalently attached to

The present sequence is the protein sequence of recombinant gelatin-like protein Hu-3. This is a tetramer of human recombinant gelatin-like protein Hu-1 (ADM48390). A claimed composition suitable as a substitute for plasma comprises a solution of saline and a protein having a colloid osmotic function. The protein is a recombinant gelatin-like protein with a molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The recombinant gelatin-like protein, or a dimer, trimer or tetramer of the protein, is useful as a plasma expander that has a lower clearance rate from blood circulation, provides better and predictable regulation of clearance rate and which is less susceptible to proteolytic degradation than presently used gelatin derivatives. Recombinant gelatin-like proteins that are in essence free of hydroxyproline do not give rise to an immunological reaction with blood samples containing IgG antibodies. The gelatin-like proteins can be covalently attached to

CC pharmaceutically active compounds. After administration, the coupled  
 CC medicament will not diffuse from the circulating blood into the  
 CC interstitium. Clearance by liver and kidney will be kept to a minimum,  
 CC ensuring a more constant plasma level of the medicament. Suitable  
 CC medicaments include those involved in intervening blood clotting,  
 CC vasodilation, function of erythrocytes, thrombocytes and leukocytes,  
 CC immune responses, and blood levels of messenger molecules such as  
 CC hormones.  
 XX  
 XX  
 SQ Sequence 821 AA;  
 Query Match 100.0%; Score 4540; DB 8; Length 821;  
 Best Local Similarity 100.0%; Pred. No. 3e-278;  
 Matches 821; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GPPGEGPTGLPGPGERGCGSRGFPAGDGVAGKGPAGRGSGPAGPKGSGCEAGR 60  
 Db |||||  
 1 GPPGEGPTGLPGPGERGCGSRGFPAGDGVAGKGPAGRGSGPAGPKGSGCEAGR 60  
 QY 61 GEAGLPGAKGLTSGSPGDPGKTGPPGAGQDGRPGPPGARGAGVGMFGPKGAA 120  
 Db |||||  
 61 GEAGLPGAKGLTSGSPGDPGKTGPPGAGQDGRPGPPGARGAGVGMFGPKGAA 120  
 QY 121 GPPGKAGRGVPPGAVGPPGAGKGEAGAGQPPGPPGAGRGSGPAGSGPGLPGPA 180  
 Db |||||  
 121 GPPGKAGRGVPPGAVGPPGAGKGEAGAGQPPGPPGAGRGSGPAGSGPGLPGPA 180  
 QY 181 GPPGKAGRGVPPGAVGPPGAGKGEAGAGQPPGPPGAGRGSGPAGSGPGLPGPA 240  
 Db |||||  
 181 GPPGKAGRGVPPGAVGPPGAGKGEAGAGQPPGPPGAGRGSGPAGSGPGLPGPA 240  
 QY 241 GPAGRGSGPAGPKGSGPCEAGRPGAGLPGAKGLTSGSPGDPGKTGPPGAGQDGR 300  
 Db |||||  
 241 GPAGRGSGPAGPKGSGPCEAGRPGAGLPGAKGLTSGSPGDPGKTGPPGAGQDGR 300  
 QY 301 GPPGPPGARGAGVGMFGPKGAGEPKKAGRGVPPGPPGAGVAGKDGAGAGQPPGPA 360  
 Db |||||  
 301 GPPGPPGARGAGVGMFGPKGAGEPKKAGRGVPPGPPGAGVAGKDGAGAGQPPGPA 360  
 QY 361 GPAGRGSGPAGSGPGLPGPAGPGEAGKGEQVPPGDLGAPGSGPAGSGPGL 420  
 Db |||||  
 361 GPAGRGSGPAGSGPGLPGPAGPGEAGKGEQVPPGDLGAPGSGPAGSGPGL 420  
 QY 421 GPPGEGCGSGRPFPCADGVAGKGPAGRGSGPAGKGEQVPPGDLGAPGSGPAGSGPGL 480  
 Db |||||  
 421 GPPGEGCGSGRPFPCADGVAGKGPAGRGSGPAGKGEQVPPGDLGAPGSGPAGSGPGL 480  
 QY 481 GSPGSGPDPGKTGPPGAGQDGRPGPPGARGAGVGMFGPKGAGPGRKAGRGVP 540  
 Db |||||  
 481 GSPGSGPDPGKTGPPGAGQDGRPGPPGARGAGVGMFGPKGAGPGRKAGRGVP 540  
 QY 541 GPPGAVGPPGAGKDGAGAGQPPGPPGAGRGSGPAGSGPGLPGPAGPGEAGKGEQ 600  
 Db |||||  
 541 GPPGAVGPPGAGKDGAGAGQPPGPPGAGRGSGPAGSGPGLPGPAGPGEAGKGEQ 600  
 QY 601 GVPDGLGAPGSGPAGSGPGLPGPGERGCGSRGFPAGDGVAGKGPAGRGSGP 660  
 Db |||||  
 601 GVPDGLGAPGSGPAGSGPGLPGPGERGCGSRGFPAGDGVAGKGPAGRGSGP 660  
 QY 661 GPKGSGPAGSGPAGLPGAKGLTSGSPGDPGKTGPPGAGQDGRPGPPGARGAGQ 720  
 Db |||||  
 661 GPKGSGPAGSGPAGLPGAKGLTSGSPGDPGKTGPPGAGQDGRPGPPGARGAGQ 720  
 QY 721 GVMGFPKGAAGPGRKAGRGVPPGAVGPPGAGKGEAGAGQPPGPPGAGRGSGP 780  
 Db |||||  
 721 GVMGFPKGAAGPGRKAGRGVPPGAVGPPGAGKGEAGAGQPPGPPGAGRGSGP 780  
 QY 781 GSPGFGQLPGPAGPGEAGKGEQVPPGDLGAPGSGPAGG 821  
 Db |||||  
 781 GSPGFGQLPGPAGPGEAGKGEQVPPGDLGAPGSGPAGG 821

ADMA48391  
 ID ADM48391 standard; protein; 617 AA.  
 XX  
 XX ADM48391;  
 XX  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 XX Human recombinant gelatin-like polypeptide Hu-3.  
 DE Plasma substitute; Gelatin-like protein; plasma expander; human.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX EP1398324-A1.  
 PN  
 XX  
 XX 17-MAR-2004.  
 PD  
 XX  
 XX 11-SEP-2002; 2002EP-00078745.  
 PF  
 XX  
 XX 11-SEP-2002; 2002EP-00078745.  
 PR  
 XX  
 XX (PUUF ) FUJI PHOTO FILM BV.  
 PA  
 XX  
 XX Bouwstra JB, Toda Y;  
 PI  
 XX  
 XX WPI; 2004-229415/22.  
 DR  
 XX  
 XX  
 PT Composition useful as substitute for plasma, comprises solution of saline  
 PT and recombinant gelatin-like protein having colloid osmotic function.  
 XX  
 PS Example 1; SEQ ID NO 2; 31bp; English.  
 XX  
 CC The present sequence is the protein sequence of recombinant gelatin-like  
 CC protein Hu-3. This is a trimer of human recombinant gelatin-like protein  
 CC Hu-1 ADM48390. A claimed composition suitable as a substitute for plasma  
 CC comprises a solution of saline and a protein having a colloid osmotic  
 CC function. The protein is a recombinant gelatin-like protein with a  
 CC molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric  
 CC point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The  
 CC recombinant gelatin-like protein, or a dimer, trimer or tetramer of the  
 CC protein, is useful as a plasma expander that has a lower clearance rate  
 CC from blood circulation, provides better and predictable regulation of  
 CC clearance rate and which is less susceptible to proteolytic degradation  
 CC than presently used gelatin derivatives. Recombinant gelatin-like  
 CC proteins that are in essence free of hydroxyproline do not give rise to  
 CC an immunological reaction with blood samples containing IGE antibodies.  
 CC The gelatin-like proteins can be covalently attached to pharmaceutical  
 CC active compounds. After administration, the coupled medicament will not  
 CC diffuse from the circulating blood into the interstitium. Clearance by  
 CC liver and kidney will be kept to a minimum, ensuring a more constant  
 CC plasma level of the medicament. Suitable medicaments include those  
 CC involved in intervening blood clotting, vasodilation, function of  
 CC erythrocytes, thrombocytes and leukocytes, immune responses, and blood  
 CC levels of messenger molecules such as hormones.  
 XX  
 SQ Sequence 617 AA;  
 Query Match 75.0%; Score 3482; DB 8; Length 617;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-207;  
 Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GPPGEGPTGLPGPGERGCGSRGFPAGDGVAGKGPAGRGSGPAGPKGSGCEAGR 60  
 Db |||||  
 1 GPPGEGPTGLPGPGERGCGSRGFPAGDGVAGKGPAGRGSGPAGPKGSGCEAGR 60  
 QY 61 GEAGLPGAKGLTSGSPGDPGKTGPPGAGQDGRPGPPGARGAGVGMFGPKGAA 120  
 Db |||||  
 61 GEAGLPGAKGLTSGSPGDPGKTGPPGAGQDGRPGPPGARGAGVGMFGPKGAA 120  
 QY 121 GPPGKAGRGVPPGAVGPPGAGKGEAGAGQPPGPPGAGRGSGPAGSGPGLPGPA 180  
 Db |||||  
 121 GPPGKAGRGVPPGAVGPPGAGKGEAGAGQPPGPPGAGRGSGPAGSGPGLPGPA 180  
 QY 121 GPPGKAGRGVPPGAVGPPGAGKGEAGAGQPPGPPGAGRGSGPAGSGPGLPGPA 180  
 Db |||||



XX DT 10-AUG-2001 (first entry)  
 XX DE Porcine alpha1(I) collagen.  
 XX KW Porcine; alpha(I) collagen; gelatin; cytostatic; viral infection;  
 KW pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;  
 KW medical; arterial sealant; bone graft; dermal implant; cancer;  
 KW rheumatoid arthritis; beverage; photographic application.  
 XX OS Sus scrofa.  
 XX FH Key Location/Qualifiers  
 FT Misc-difference 929..830  
 FT /note= "Encoded by ggcgaacctggtgatgctgctgtaaggcgtat  
 FT ctggtccccccgcccctgctgga"  
 XX WO2001134647-A2.  
 XX PD 17-MAY-2001.  
 XX PF 10-NOV-2000; 2000WO-US030792.  
 XX PR 12-NOV-1999; 99US-00439058.  
 PR 10-NOV-2000; 2000US-00709700.  
 XX (FIBR-) FIBROGEN INC.  
 XX PI Bell MP, Neff TB, Polarex JW, Seeley TW;  
 XX WIPI; 2001-335911/35.  
 DR N-PSDB; AAD06576.  
 XX PT Novel isolated and purified bovine or porcine collagens and gelatins  
 PT useful in medical, pharmaceutical, food and cosmetic industries, as  
 PT vaccine, and for treating autoimmune disorders, infections and cancer.  
 XX Example 3; Fig 8; 168pp; English.  
 CC The present sequence is porcine alpha1(I) collagen. The present invention  
 CC relates to recombinant synthesis of collagens and gelatins derived from  
 CC animals. Collagen is useful in medical, pharmaceutical, food and cosmetic  
 CC industries. Collagen is an important component of arterial sealants, bone  
 CC grafts, drug delivery system, dermal implants, haemostats, and  
 CC incontinence implants, and for treating autoimmune disorders such as  
 CC rheumatoid arthritis. Collagen is useful in food products such as sausage  
 CC casings, and in cosmetics or facial and skin products such as  
 CC moisturisers. Recombinant gelatin is useful in vaccine formulations for  
 CC treating viral infections, autoimmune diseases and cancer. Gelatin is  
 CC useful in the manufacture or as a component of various pharmaceutical and  
 CC medical devices and products, in food and beverage industries, in hair  
 CC care and skin care products, as a glue or adhesive in various  
 CC manufacturing processes, as a light-sensitive coating in various  
 CC electronic devices, as photoresist base in photolithographic processes,  
 CC in printing and photographic applications, in laboratory application, and  
 CC as a component in various gels used for biochemical and electrophoretic  
 CC analysis, including enzymographic gels  
 XX Sequence 1449 AA;  
 Query Match 59.6%; Score 2765.5; DB 4; Length 1449;  
 Best Local Similarity 56.7%; Pred. No. 1.4e-162;  
 Matches 559; Conservative 32; Mismatches 226; Indels 169; Gaps 21;  
 QY 1 GPPGPGPTGLPSPGRRGFGS-----RGFFCA-----DGVAGPKGPAGRGSPGPA 48  
 Db 193 GPPGAPGFGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 252  
 QY 49 GPKSGPGGAGRP-----GEAGLPGAKGLTSPGSPGPDGKTGPPGAGQD 93  
 Db 253 GARGLPATGLPMKMGHRSGLDAGKADGAPAGPKGPGSGENAPGQMGRPLPGR 312  
 QY 94 GRGPGPGPGARGQ---AGVMFGPGPKGAAGE---PKAGRGVPPGAVGAPAGKDGGA 147

Db 313 GRGPGPGAGAGNDGATGAAGPPGPTGAPGPPGPAVGAKGEAGPQARGSEGPQGV 372  
 QY 148 GAQGPAGPAGAGERGEQGP-----AGSGFQGLPGPAGPPGEAKGPEQ 192  
 Db 373 GEPGPPGAGAGPAGNPGADQPGKKGANGAPGAGPFGFARGSPGSPGPPGPK 432  
 QY 193 GVEGDLGAP-----GPSGPA-----GEPGPTCLPAPPGERGQ 225  
 Db 433 GNSGEPGAPSGKDTGAKGEPGPTGVQGPFGPAGEEKGARGEPGAPGLPAPPGERG 492  
 QY 226 GSRGPGADGAVAGPKGPAGERSGPPAGPKGSPGAGRPGBAGLPGCAKGLTSPGSP 285  
 Db 493 GSRGPGADGAVAGPKGPAGERSGPPAGPKGSPGAGRPGBAGLPGCAKGLTSPGSP 552  
 QY 286 GKTGPPGAGQDGRPGPPGPGARGOAGVMGPPGKGAAGBPKAGRGVGPFGPAGVGA 345  
 Db 553 GKTGPPGAGQDGRPGPPGPGARGOAGVMGPPGKGAAGBPKAGRGVGPFGPAGVGA 612  
 QY 346 GKDGAGAGQPPGPGPAGERGEQGPAGSPGQGLPGPAGPPGEAKGPEGVEPGLGAP 405  
 Db 613 GKDGAGAGQPPGPGPAGERGEQGPAGSPGQGLPGPAGPPGEAKGPEGVEPGLGAP 672  
 QY 406 GPSGPAGE-----PGP-----TGLPGPPGSGGSGSRGFP 435  
 Db 673 GFSGARGRGFPGERGVQPPGPGPAGPRGANGAPNDGAKGDAGAPGAPSGOAPGLQMP 732  
 QY 436 GADGAVAGPKGPAGERSGPPAGPKGSPGE-----AGRPGEAGLPGCAKGLTSPGSP 486  
 Db 733 GERGAAGLPKPKGDRDAGPKGADGAPKDGVRGLTGPIGPPGAPAGPDKGETGSPGA 792  
 QY 487 GPDGKTGPPGAGQDGRPGP---PPPGARGOAGVMGF---PGPKGAAGBPKAGRGV 539  
 Db 793 GTGARGAPDRGEPGPPGPGPAGFAGPPGADGQPGKAGGPTGPPGIVGVPAGPKGARG 852  
 QY 540 PPGPCA-----VGPAGKDXGAGAOQPPGPGAPAGE---RGEQGPAGSPGQGLPGP 587  
 Db 853 AGPPGATGPPGAAGRVGPPGPGNAGPPGPGAGKSGKGRGETGAPRGAGEAGPPG 912  
 QY 588 AGPPEAKGPEBQVPGDLGAPGSPGAPGEPGPTGLPDPGGERGPGSRGPPGADGAVG 647  
 Db 913 PGPAGEKSPGADGAPAGPPTGPGQIAGQGVVGLPQQRGERGPPGLPGSPGSPGQGP 972  
 QY 648 KPAGERGSPGAPKGPAGRGAPGEPGAGLPGCAKGLTSP---GSPGD---GKTGPPG 701  
 Db 973 SGPSGERGPPGMPG---PGLAGPPGSGRGAGSGPDRDAPGPKGDRGSGPAGP 1029  
 QY 702 AGQDGRPGPPGARGOAGVMGPPGKGAAGBPKAGRGVGPFGPAGVAGPKDGE--- 758  
 Db 1030 PGAPGAPGAPGVPAGKSGDRGETGAPGAPGVPVGPARGPAGPQGRGDKGTGEGQD 1099  
 QY 759 -----ACAQPPGPG-----AGPAGERGQGPAGSP---GFGGLPGPAGP 794  
 Db 1090 RGKIHRSFGLQPPGPPGPGSPGQSGSAGSPAGRPGPPGAGAPKDGKGLNGLPPTGP 1149  
 QY 795 PGEAGKPGEQGVPGDLGAPGSPGAG 820  
 Db 1150 PPRGRTGDAGVPVGPFGPPGPPGPPG 1175

RESULT 5  
 AAY84544  
 ID AAY84544 standard; protein; 1057 AA.  
 XX  
 AC AAY84544;  
 XX  
 DT 25-JUL-2000 (first entry)  
 XX  
 DE A human collagen 1 (alpha) protein helical region.  
 XX  
 KW Extracellular matrix protein; self aggregation; hydroxylated proline;  
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;  
 KW collagen; fibrinogen; fibronectin; post translational hydroxylation.



PA (BUEC/) BUECHTER D D.  
 PI Paoella DN, Gruskin EA, Buechter DD;  
 XX WPI; 2000-271051/23.  
 DR N-PSDB; AA299843.  
 XX  
 PT Incorporating non-natural amino acid into polypeptide, useful e.g. for  
 PT production of bioadhesives, by epoxidation or substitution of  
 PT dehydroproline residues.  
 XX  
 PS Disclosure; Fig 6; 66pp; English.  
 XX  
 CC The present sequence represents a human type 1 (alpha1) collagen protein.  
 CC Peptides derived from the protein were used to demonstrate incorporation  
 CC of 3,4-dehydro-L-proline into the peptide, using the method of the  
 CC invention. The specification describes a method for the incorporation of  
 CC non-natural amino acid into a polypeptide. The method comprises reacting  
 CC at least one 3,4-dehydroproline residue in the polypeptide with an  
 CC epoxidation reagent from a polypeptide containing at least one 3,4-  
 CC epoxyproline residue. The method is used for studying the effects of non-  
 CC natural amino acids on structure and function of polypeptides. The method  
 CC is also useful for commercial production of collagen or mussel adhesive  
 CC proteins (which are useful as bioadhesives), and for incorporating a wide  
 CC variety of groups, including therapeutic ligands and biological probes,  
 CC into polypeptides.  
 XX  
 SQ Sequence 1058 AA;

Query Match  
 Best Local Similarity 59.6%; Score 2765; DB 3; Length 1058;  
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGPGPTGLPFPGERGSGSGFGADGVAGKPGAGRGSPGAGPKGSGPEA--- 57  
 Db 19 GPMGSGPRGLPFPFGAGPQGFQGPGEPCGASGPMGPRGPPGPKNGDDGEAGKP 78  
 QY 58 GPGEAGLP---GAKLTGSGSGP-----DKTGTTPAGQDGRPPGPPGAP 105  
 Db 79 GRPGERGPPGQAGRLPGTGLGKMGHRSGLDGAAGKADGAPGKGPSPGAGAP 138  
 QY 106 QGAVMGPPGPKGAAGEPKAGER-----GVPPGPAVGA-----GKDGGAQ 150  
 Db 139 QMGPRGLPGERGPRGAPGAGAGNDGATGAAGPPTGAPPPGPAVGAAGKAGAPQ 198  
 QY 151 GPPGAGPAGRGSGP-----GPPGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
 Db 199 GPRGSEGGVGRGPPGPPGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 258  
 QY 181 GPPGAGKPGGQGVPGDLGAP-----GPPGAGAGAGAGAGAGAGAGAGAG 213  
 Db 259 GPQGPAGPAGKNGSGPAGSGKDTGAKGEPGVGVGPPGAGBEGKRGAGGEPGT 318  
 QY 214 GLPFPGERGPGSGRFPFGADGVAGKPGAGRGSGPGAGPKGSPGAGRPGEAGLPAG 273  
 Db 319 GLPFPGERGPGSGRFPFGADGVAGKPGAGRGSGPGAGPKGSPGAGRPGEAGLPAG 378  
 QY 274 GLTSGSGPDGKTGTPGAGQDGRPPGAGQAGVWGFPDPKGAAGFPAGKAGER 333  
 Db 379 GLTSGSGPDGKTGTPGAGQDGRPPGAGQAGVWGFPDPKGAAGFPAGKAGER 438  
 QY 334 GVPFPAGVAGPAGKDGAGAGQGGPPGAGAGRGSGPGFGLPAGPAGPAGKAGP 393  
 Db 439 GVPFPAGVAGPAGKDGAGAGQGGPPGAGAGRGSGPGFGLPAGPAGPAGKAGP 498  
 QY 394 GEQGVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 423  
 Db 499 GEQGVPGDLGAPGSGARGRFPGERGVQPPGAGPRGANGAGNDGAKGADGAPGAP 558  
 QY 424 GERGGPSRGPAGADGVAGKPGAGRGSGPGAGPKGSPCE-----AGSPGAGLP 474  
 Db 559 GSQAGFLQGMFGERGAGLPFGKGRDGDGKADGSKDGKGRGLTGPFGPAGAP 618

QY 475 GAK-----GLTSGSGSPGDKTGTTPGAGQDGRPPGPPGAGQ----- 515  
 Db 619 GDKGSGSPGAGTGTGAGAGPDRGPPGPPGAGFAGCAGQCPGAKGPGDAGAKGDA 678  
 QY 516 -----AGVVGFPKGAAGEPKAGRGVPGPPGA-----VGPAGKDGAGAGQGP 561  
 Db 679 GPPGAPGAGPPGPIGNVAGAPGAKGARGAGPPGATGFFGAAGRVGPPGSGNAGPPGP 738  
 QY 562 GPAGP---AGEGEGPAGSPGFGGLPAGPAGPAGGEGQGVPGDLGAPGSPGAPGP 618  
 Db 739 GPAGKEGKGPRGTGTGAGRGVEVPGPPGAGKSGPGADGAPGATPGPQGIAGOR 798  
 QY 619 GPTGLPFPGERGSGRFPFGADGVAGKPGAGRGSGPAGP---KSPGAGRPGEA 675  
 Db 799 GVVGLPQGRGEGFPGLPFGSGEPGKQSGASGERGPPGPMGPPGLAGPPGESREGAP 858  
 QY 676 GLPFAKGLTSGSGSPGPDGKTGTTPGAGQDGRPPGPPGAGRGQAGVWGFPDPKGAAGEP 735  
 Db 859 GABGSGPRDSSPGAKGDRGTGTGAPGAPGAPGVPVGPAGKSGDRGTGTGAPGAPV 918  
 QY 736 GKAGRGVPGPPGAVGAPGAGKDE-----AGAQQPPGP-----AGPA 771  
 Db 919 GPAGARGPAGPQGRDKGTGTGAGRGVGLGHRGSLGQGGPPGPPGSGEQSGASGPA 978  
 QY 772 GERGEGPAGSP---GFOGLPAGPAGPAGKPGEGVPGDLGAPGSPGAP 820  
 Db 979 GPRGPPGAGAGKDLNGLPPIGPPGPRGTGAGVVGPPGPPGPPG 1030

RESULT 7  
 AA584540  
 ID AAY84540 standard; protein; 1107 AA.  
 XX  
 AC AAY84540;  
 XX  
 DT 25-JUL-2000 (first entry)  
 XX  
 DE Amino acid sequence of a chimeric collagen 1 (alpha1)/decorin protein.  
 XX  
 KW Extracellular matrix protein; self aggregation; hydroxylated proline;  
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;  
 KW collagen; fibronectin; fibronectin; post translational hydroxylation;  
 KW decorin; chimera.  
 XX  
 OS Homo sapiens.  
 OS Unidentified.  
 OS Chimeric.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 858  
 FT /note= "Gly encoded by GCT"  
 XX  
 PN EP992586-A2.  
 PD 12-APR-2000.  
 XX  
 PF 07-OCT-1999; 99EP-00119184.  
 XX  
 PR 09-OCT-1998; 98US-00169768.  
 XX  
 PA (USSU) US SURGICAL CORP.  
 XX  
 PI Gruskin EA, Buechter DD, Zhang G, Connolly K;  
 XX  
 DR WPI; 2000-259138/23.  
 DR N-PSDB; AAA12500.  
 XX  
 PT Production of extracellular matrix proteins containing 4-trans-  
 PT hydroxyproline results in native self aggregating proteins, useful on  
 PT medical implants.  
 XX  
 PS Claim 24; Fig 18; 260pp; English.  
 XX

The specification describes a method for producing an extracellular matrix protein or its fragment. The extracellular matrix protein is capable of self aggregating in a cell which does not ordinarily hydroxylate prolines. The method comprises optimising a nucleic acid sequence for expression in the cell by substitution of codons preferred by that cell for naturally occurring codons not preferred by the cell; incorporating the nucleic acid sequence into the cell; and contacting the cell with a hyperosmotic growth medium containing at least one amino acid, selected from the group consisting of trans-4-hydroxyproline and 3-hydroxyproline to allow at least one of the amino acids to be assimilated into the cell and incorporated into the extracellular matrix protein. The method may be used to make host cells assimilate and incorporate trans-4-hydroxyproline into proteins. This is especially useful in the recombinant production of proteins such as collagen, fibrinogen and fibronectin whose ability to self aggregate and produce functional proteins depends on the post translational hydroxylation of proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans-4-hydroxyproline. The present sequence represents a chimeric collagen 1 (alpha1)(decorin protein, which may be produced using the method of the invention

Query Match	59.68;	Score 2765;	DB 3;	Length 1107;
Best Local Similarity	55.44;	Pred. No. 1.2e-128;		
Matches 561;	Conservative 31;	Mismatches 232;	Indels 192;	Gaps 19;
Qy	1	GPGECPGTGLPGPPGRRGGPSGRGFCADGVAGPKGPAGBERGSPGAGPKGSGEAA---	57	
Db	18	GPVGPSPRGLPGPPGAFGPGCGFPGEGEPGASGPMGRGPPGPKNGDDGEAGK	77	
Qy	58	GRGGEAGLP---GAKGLTGSFGSGP-----DQKTGPPGAGQDGRPPGPPGAG	105	
Db	78	GRPCRGPPOGARGLPGTAGLPMKMGHRGSLDGAAGDAGPAGPKGEPGSGAGAP	137	
Qy	106	GOAGVFMFPKGAAGPRKAGER---GVPPPGAVGPA-----GKDGEAGAG	150	
Db	138	GQMGRGLPBERGPFAGPAGARGNDGATGAAGPGTGPAGPPGPAVKAKEAGGP	197	
Qy	151	GPPGPAGPAGBERGQGP-----AGSFGQGLPGPA	180	
Db	198	GPRSEGPQVGRPPGPPGAGAGPAGNPAGDQPGAKGANGAPGIAGAFGFGARGFS	257	
Qy	181	GPPGEAGKPGSQGVPGDLGAP-----GPGGPA-----GEPPT	213	
Db	258	GFQPGGPPGPKGNSGEBPCAPGSKGDTCAKGEPPGVQGGPPGAGBERGKARGEPPT	317	
Qy	214	GLPDPBERGCGPSRGFFGADGVAGPKGPAGERSGPGAPKSGPGEAGRPGEAGLPGAK	273	
Db	318	GLPDPBERGCGPSRGFFGADGVAGPKGPAGERSGPGAPKSGPGEAGRPGEAGLPGAK	377	
Qy	274	GLTSGPSGPDGKTGTPGPGAGQDGRPPPPPGARGQAGVMGFPFGAAGEPGKAGER	333	
Db	378	GLTSGPSGPDGKTGTPGPGAGQDGRPPPPPGARGQAGVMGFPFGAAGEPGKAGER	437	
Qy	334	GVPPGPVAGVGPAGKDGEGAGAGCPGPGAPGABERGQGPAGSPGFGGLPGPAGPPGEAGKP	393	
Db	438	GVPPGPVAGVGPAGKDGEGAGAGCPGPGAPGABERGQGPAGSPGFGGLPGPAGPPGEAGKP	497	
Qy	394	GEQGVPGDLGAPGPGSGPAGE-----PGP-----TGLPGPP	423	
Db	498	GEQGVPGDLGAPGPGSGAGERFPBERGVQGGPPGAPRFGANGAPGNDCAKDKDAGAPAG	557	
Qy	424	BERGCGPSRGFFGADGVAGPKGPAGERSGPGAPKSGPGE-----AGRPGEAGLP	474	
Db	558	GSQGFAPGLQCMFEGERGAAGLPKPKDGRDAGAPKADGSPGKDGVRGLTGPIGPPGPAGAP	617	
Qy	475	GAK-----GLTSGPSGSPDGKTGTPGAGQDGRPPGPPGAGRCQ-----	515	
Db	618	GDKGESGSPGAPGTGARGAPDGRGPPGPPGAGPAGPPGADQGPAGKEPGDAGAKGDA	677	
Qy	516	-----AGVMGFFGPKGAAGEPKKAGERGVPGPPGA-----VGPAGKDGEGAGAGQGP	561	





171-1454 of the protein and around 500 bp of the 3' untranslated region.  
The 2 clones were used to generate a number of fragments which were used  
to construct composite sequences encoding variant collagen molecules. The  
fragments are: (A) containing nucleotides (nt) -4 to 479; (B) containing  
TAA upstream of the sequence encoding the PRS (pathogenesis-related  
protein S) signal peptide and bases 66-77 from the sequence encoding the  
N-terminus of the pro-collagen amino propeptide domain; (C) the whole of  
the amino propeptide domain (nt 72-479); (D) all of the amino-telopeptide  
domain (nt 474-534) and the N-terminus of the helical region (nt 535-  
1920); (E) the DRAII-BamHI fragment (1709-2808) of alpha22, encoding aa  
567-936 of the central helical domain; (F) the BamHI-EcoRI (2803-4362)  
region of alpha22, encoding aa 936-1192 in the central helical domain  
and aa 1193-1454 in the C-propeptide domain; (G) the C-terminus of the C-  
propeptide domain (aa 1346-1464) plus stop codons, and (H) as G but  
encoding aa 1343-1401 and also including the KDEL motif for retention in  
the ER. This sequence represents a recombinant human collagen. The  
encoding gene was constructed from fragments (A), (D), (E), (F) and (G).  
The recombinant gene is used for expression of mammalian collagen in  
plant cells. The transformed plants, their extracts and parts are useful  
as biomaterials (haemostatic compresses, sponges or bandages) and in  
pharmaceutical, medical, odontological, cosmetic and biotechnological  
compositions (e.g. as prostheses for cardiac valves, ligaments or tendons  
; skin substitutes; gingival implants; microcapsules for perfumes; guide  
tubes for nerve regeneration; slow release products for antibiotics,  
growth factors, anticancer agents or anti-inflammatories; surgical thread  
and components of ointments). They are suitable for treating any disorder  
related to collagen dysfunction and gelatin, produced from collagen, is  
used to produce glues, surgical prostheses and foods

Query Match 59.6%; Score 2765; DB 2; Length 1464;  
Best Local Similarity 55.4%; Pred. No. 1.5e-162;  
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGPGTGLPGRGGRGSRGFFGADGVAGPKGAPGERGSGFPAGPKGSPCEA--- 57  
DB 179 GPMGSGRGLGPPGAGPQGFQFPCEGPGASGPMGPRGPPGKNGDDGEAGK 238  
QY 58 GRPGAGLP---GAKGLTSGSPGCP-----DGKTPGAGQDGRGPPGPGAR 105  
DB 239 GRPGRGPPGPGQAGRLFTAGLPMKGRHFGSLDGAKGADGAPAGPKGSGPENGAP 298  
QY 106 GQAGVMGPPGPKGAGEPKAGER-----GVGPPGAVGPA-----GKDGAGAG 150  
DB 299 GQMGFRGLGGRGRGAPGAPAGARGNDGNTAGPFGTGTGAGPFGFVAVGAKGAGPQ 358  
QY 151 GPPGAPGAPAGERGEGP-----AGSPGFQGLPQPA 180  
DB 359 GPRGSEGPQVAGEFPGPPGAGAGPAGNPGADGQKANGANGAPGAGPFGGARGPS 418  
QY 181 GPPGAGKXGEGVPGDILGAP-----GPSGPA-----GEPGPT 213  
DB 419 GPQGGPPGPPKNGSGEPGAPSGKGTGAKGPPGVGQPPGAGEGKRGARGEPPT 478  
QY 214 GLPGRGGRGSGSRGFFGADGVAGPKGAPGERGSPGAPKXGSGEAGRPGEAGLPKAK 273  
DB 479 GLPGRGGRGSGSRGFFGADGVAGPKGAPGERGSPGAPKXGSGEAGRPGEAGLPKAK 538  
QY 274 GLTSGSGSGPDGKTGPPGAGQDGRPPGPPGARGQAGVMGFFGPKGAGEPKKAGER 333  
DB 539 GLTSGSGSGPDGKTGPPGAGQDGRPPGPPGARGQAGVMGFFGPKGAGEPKKAGER 598  
QY 334 GVPGPPGAVGPKGDEAGAQPPGAPGAPGERGEGPAGSPGFQGLPAGPAPGGEAGKP 393  
DB 599 GVPGPPGAVGPKGDEAGAQPPGAPGAPGERGEGPAGSPGFQGLPAGPAPGGEAGKP 658  
QY 394 GEGVFGDLCAPGSPGAGE-----PGP-----TGLPFP 423  
DB 659 GEGVFGDLCAPGSPGARGERPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 718  
QY 424 GERGGSGRGFFGADGVAGPKGAPGERGSGPAGPKGSPGE-----AGRPGEAGLP 474

DB 719 GSQAPGLQGMGERGAAGLPGPKDRGDGAPKGDGADGSPKGDVRLTGLTPIGPPGAP 778  
QY 475 GAK-----GLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQ----- 515  
DB 779 GDKGESGSPGAPGTPGARGAPDRGEPGPPGAPGAPGPPGADGQPKAGXGEPGDAGAKGDA 838  
QY 516 -----AGVMGFFGPKGAGEPKKAGERGVGPPGA-----YGPAGKDGCEAGAQGPP 561  
DB 839 GPPGAPGAPGPPGPIGNVAPAKGARGAGAGPFGATGTPGAAGRVPGPSNAGPFP 898  
QY 562 GPAGP---AGERGEGPAGSPGFQGLPAGPFPGEAGKPGEGVPGDLGAPGSPGAP 618  
DB 899 GPAGKGGKGRGTGAPAGPGEVGPFPAGGKSGPAGADGAPAGPCTGPGQIAGOR 958  
QY 619 GTGLPGRGGRGSRGFFGADGVAGPKGAPGERGSGPAGP---KSGPCEAGRPGEA 675  
DB 959 GVVGLPQGRGGRGFFGLPAGPGEKQSGASGSRGPPGMPGPPCLAGPPGSGREGAP 1018  
QY 676 GLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGAPGARGQAGVMGFFGPKGAG 735  
DB 1019 GAGSGRGSPPGAKGDRGTGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 1078  
QY 736 GKAGERGVGPPGAVGAPGAKDGE-----AGAQGPPG-----AGPA 771  
DB 1079 GPAGARGPAGPQGRGDKGTGEGQDRGKIKHGRGSLGQPPGPPGSGPGEQSGSAG 1138  
QY 772 GERGEGPAGSP---CFQGLPAGPAGPGEAGKPGEGVPGDLGAPGSPGAP 820  
DB 1139 GPRGPPGSAAGPKGKGLNGLPFIIPGPPRGRTGADGAPGPPGPPGPPG 1190

RESULT 10  
ADD45059  
ID ADD45059 standard; protein; 1464 AA.  
XX  
AC ADD45059;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human Protein P02452, SEQ ID NO 10491.  
XX  
KW Human; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX  
OS Homo sapiens.  
XX  
PN W02003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
PA (GEHO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
XX WPI; 2003-268312/26.  
DR GENBANK; P02452.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also

CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1464 AA;

Query Match 59.6%; Score 2765; DB 7; Length 1464;  
 Best Local Similarity 55.4%; Pred. No. 1.5e-162;  
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGPGTGLPGRGGRGSRFFPGADGVAGPKGAGRGSPGPGAGKSGPGEA--- 57  
 DB 179 GPMFSGPRGLPFGPAGPQGFQPGCEPGEPCASGPMWRGPPPPGKNGDDGEAGKP 238  
 QY 58 GEPGAGLP---GAKGLTSGSPGSP-----DGKTGPPGAGDGRPPGPPGAR 105  
 DB 239 GRPGERGPPGQARGLPCTAGLPQMKGHRGFLDGAAGKAGDAGPAGPKGEPGSGENGAP 298  
 QY 106 GQAGVMGPPGPKGAAGEPKGAGER-----GVPGPAGVGA-----GKDGAGAG 150  
 DB 299 GQMGPRGLPGERGRGAGPAGAGCNDGATCAAGPPGTGAPGPPGAVGAKGEAGPQ 358  
 QY 151 GPPGAGPAGRGEGGP-----AGSPFGQGLPGPA 180  
 DB 359 GPRGSEPGQVRGEPGPPGAGAGPAGNPGADGQPGAKGANGAPGIAGAPGPPGARGPS 418  
 QY 181 GPPGAGKPGQGVPGDLGAP-----GPSGA-----GEPGPT 213  
 DB 419 GPGGPGPPGKNGSEPCAGSGKDTGAKGEPGFVQGGPPGABEGKRGARGEPGPT 478  
 QY 214 GLPGPGRGGRGSRGFPDGVAGPKGAPAGERSPGPAGPKSGPAGRPGGAGLPGAK 273  
 DB 479 GLPGPGRGGRGSRGFPDGVAGPKGAPAGERSPGPAGPKSGPAGRPGGAGLPGAK 538  
 QY 274 GLTSGSPGPDGKTGPPRAGQGRPPGPPGARGQAGVMGPPGKGAAGSPGKAGER 333  
 DB 539 GLTSGSPGPDGKTGPPRAGQGRPPGPPGARGQAGVMGPPGKGAAGSPGKAGER 598  
 QY 334 GVPGPAGVAGPKDGEAGAGPFGPAGPAGERGEGQGPAGSPFGQLPGPAGPGEAGKP 393  
 DB 599 GVPGPAGVAGPKDGEAGAGPFGPAGPAGERGEGQGPAGSPFGQLPGPAGPGEAGKP 658  
 QY 394 GEGVPGDLGAPSGPAGE-----PGP-----TGLPGPP 423  
 DB 659 GEGVPGDLGAPSGPAGE-----PGP-----TGLPGPP 423  
 QY 424 GERGGPSRFPAGDGVAGPKGAPAGERSPGPAGPKSGPGE-----AGRPGGAGLP 474  
 DB 719 GSQAGPLQGMFGRGAAGLPKPKGDRGADGAPKAGDSFKDGVRLTGPFGPPGAGLP 778  
 QY 475 GAK-----GLTSGSPGPDGKTGPPRAGQGRPPGPPGARGQ----- 515

DB 779 GDKGESGSPGAGTGAAGAPGDRGPPGPPGAGFAGPFGADGQPGAKGPGDAGAKGDA 838  
 QY 516 -----AGVMGFPKPKGAAGEBPGKAGRGVFGPPPCA-----VGPAGKDGEGAGQGGP 561  
 DB 839 GPPGAGPAGPPGPIGNVGAFAKAGAGAGPPGATGFFCAAGRVGPPGSGNAGPPGPP 898  
 QY 562 GPAGP---AGERGEGPAGSPGQGLPGPAGPPEAGKPEGQGVPGDLGAPGSPGAGP 618  
 DB 899 GPAGKEGKGRGETGAPRGPEGVGPPGPPGAGEKSGPCADGACAPAGTPTGPGIAGQR 958  
 QY 619 GPTGLPGRGGRGSPGSRGFPAGDGVAGPKGAPAGERSGPPAGP---KSGPGEAGPPGA 675  
 DB 959 GVVGLPQQRGERFPGLPSPSEFPKQSGSGASGERGPPGPPGGLAGPPGSGREGAP 1018  
 QY 676 GLPAGKLTGSPGSPGDKTGTTPGAGQDGRPPGPPGARGQAGVMGFPKGAAGP 735  
 DB 1019 GAEGSPGRDGSAGKAGDRGETGTGAPGAPGAPGAPGVPVGPAGKSGDRGETGAPGAPV 1078  
 QY 736 GKAGERGVGPPGCAVCPAGKDG-----AGAQPPGP-----AGPA 771  
 DB 1079 GPAGARGPAGPQGRDKGETGEGQDRGKTKHHRGSLGQGGPPGPPGSGPGEQGPSGASGA 1138  
 QY 772 GERGEGPAGSP---GFGQLPGPAGPPGAGKPGGQGVGPDLCAGPSPGAP 820  
 DB 1139 GPRGPPGSAGAPGKDLNGLPGPIGPPGPRGRTGAGPVGPPGPPGPPG 1190

RESULT 11

ADD45055

ID ADD45055 standard; protein; 1464 AA.

XX AC ADD45055;

XX DT 29-JAN-2004 (first entry)

XX DE Human Protein P02452, SEQ ID NO 10487.

XX KW Human; pain; neuronal tissue; gene therapy;

XX KW spinal segmental nerve injury; chronic constriction injury; CCI;

XX KW spared nerve injury; SNI; Chung.

XX OS Homo sapiens.

XX PN WO2003016475-A2.

XX PD 27-FEB-2003.

XX PF 14-AUG-2002; 2002WO-US025765.

XX PR 14-AUG-2001; 2001US-0312147P.

XX PR 01-NOV-2001; 2001US-0346382P.

XX PR 26-NOV-2001; 2001US-0333347P.

XX PA (GEO) GEN HOSPITAL CORP.

XX PA (FARB) BAYER AG.

XX PI Woolf C, D'urso D, Befort K, Costigan M;

XX DR WPI: 2003-268312/26.

XX DR GENBANK; P02452.

PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.

XX PS Claim 1; Page; 1017pp; English.

CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent

CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC the sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1464 AA;

Query Match 59.6%; Score 2765; DB 7; Length 1464;  
 Best Local Similarity 55.4%; Pred. No. 1.5e-162;  
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGEPPTGLPGPGRGGSGFPGADQVAGPKPAGRGSPGAGPKGSPGEA--- 57  
 DB 179 GPMGSGRGLPGPAGPAGPQGGFPGEPGEPGASGPMGRGPPGKNGDDGAGKP 238  
 QY 58 GRPGEAGLP---GAKGLTSGSPGSP---DGKTGPPGAGQDGRPPGPPGAP 105  
 DB 239 GRPGERPPGQAGRLPGTAGLPMKMGHRGFSGLDGAKDAGAPGKGPSPGENGAP 298  
 QY 106 QGAVMGPPGPKGAAGRGKAGER-----GVPPPGAVGPA-----GKDGGAQA 150  
 DB 239 QCMGPRGLPGRGRPCAPGAPAGRGNDGATGAAGPPOPTGAPGPPGAVGAKEAGPQ 358  
 QY 151 GPPGAPGAGRGEGP-----AGSPFGQLPGPA 180  
 DB 359 GPRSEGGVGRGFPFGPAGAGPAGNPGADQPGKAGKANGAPGIAGAPGFGARGPS 418  
 QY 181 GPPGAGKPGQGVFDLGA-----GPSGA-----GEGPT 213  
 DB 419 GPQGGPGLPKGNSGEPGAPGSKGDTCAKGPFPVGVQPPGAGEGKRGARGEFPT 478  
 QY 214 GLPGPGRGGGSGFPGADQVAGPKPAGRGSPGAGKSGPGRGPRGPRGAGLPCA 273  
 DB 479 GLPGPGRGGGSGRFFPGADQVAGPKPAGRGSPGAGKSGPGRGPRGPRGAGLPCA 538  
 QY 274 GLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEGKAGER 333  
 DB 539 GLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEGKAGER 598  
 QY 334 GVPGPAGVAGPKDGERAGAQPPGAPGAGERGEGQAGSPGQGLPGPAGPGEAGKP 393  
 DB 599 GVPGPAGVAGPKDGERAGAQPPGAPGAGERGEGQAGSPGQGLPGPAGPGEAGKP 658  
 QY 394 GEQGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423  
 DB 659 GEQGVPGDLGAPGSPGAGERGFPGERGVQPPGAPRGANGAPGNDGAGDAGAP 718  
 QY 424 GERGGSGRGGFPGADQVAGPKPAGRGSPGAGKSGFGE-----AGRGEGALP 474  
 DB 719 GSQGAPGLQCMFGERGAAGLPKGDGADGPKGADGSGFKDGVNGLTGTPTGPPGAP 778  
 QY 475 GAK-----GLTSGSPGPDGKTGPPGAGQDGRPPGPPGAPGQ----- 515  
 DB 779 GDKGSSGGPAGPTGARGAPDRGEPGPPGAPGAPGADGQGAKGEGCDAGAKDA 838  
 QY 516 -----AGVMGFPFGPKGAAGEGPKAGERGVGPPGA-----VGPAGKGEAGQAGPP 561

DB 839 GPPGAPGAPGPPPTGNVGAFAKARGSGAGFPAGATGPPGAGRVGPPGSGNAGPPGPP 898  
 QY 562 GPAG---AGRGSGGAGSGFQGLPGPAGPPGAGKPGQGVFDLGAFAFGSPGAPGP 618  
 DB 899 GPAGKEGKGPRGETGAPRGCEVGPFPAGPKSGPADGAPAGTGTGPGQIAGQR 958  
 QY 619 GPTGLGPPGPPGSGRGGFPGADQVAGPKPAGRGSPGAPG---KGSFGEAGRPGEA 675  
 DB 959 GVVGLPGQGRGPRGFPGLPGSGEPKQKQFSGASGERGPPGPMGPPGLAGPFGESEGAP 1018  
 QY 676 GLPFAKGLTSGSPGSPGKTPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEP 735  
 DB 1019 GAESGPRGDSGPAKGRGTGTGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPV 1078  
 QY 736 KGXGGRGVGPPGPPGAVGAPGAKDGE-----AGAQQPPGP-----AGPA 771  
 DB 1079 GPAGARGPAGPQGRDGRGEGQDGRGKGRHGFSLQGGPPGPPGSGEQGPPSGASGPA 1138  
 QY 772 GERGGQAGSP---GFGQLGPPAGPGEAGKPGQGVFDLGAFAFGSPGAPG 820  
 DB 1139 GPRGPPGAGAPGKDLNGLPGFPGPRGRTGDPGVPVGPFPGPPGPPG 1190

# RESULT 12

ADD45051

ID ADD45051 standard; protein; 1464 AA.

XX AC ADD45051;

XX AC 29-JAN-2004 (first entry)

XX DE Human Protein P02452, SEQ ID NO 10483.

XX KW Human; pain; neuronal tissue; gene therapy;

XX KW spinal segmental nerve injury; chronic constriction injury; CCI;

XX KW spared nerve injury; SNI; Chung.

XX OS Homo sapiens.

XX PN WO2003016475-A2.

XX PD 27-FEB-2003.

XX PF 14-AUG-2002; 2002WO-US025765.

XX PR 14-AUG-2001; 2001US-0312147P.

XX PR 01-NOV-2001; 2001US-0346382P.

XX PR 26-NOV-2001; 2001US-0333347P.

XX PA (GEO ) GEN HOSPITAL CORP.

XX PA (FARB ) BAYER AG.

XX PA Woolf C, D'urso D, Befort K, Costigan M;

XX PI WPI; 2003-268312/26.

XX DR GENBANK; P02452.

XX PT New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

XX PS Claim 1; Page; 1017pp; English.

XX CC The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially

expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 1464 AA;

Query Match 59.6%; Score 2765; DB 7; Length 1464;  
Best Local Similarity 55.4%; Pred. No. 1.5e-162;  
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGPGTGLPGRGSGRPFPGADGAGVAGPKGPPAGRGSPGAGPKGSGEA--- 57  
DB 179 GPMGSGRGLPFPAGPFGQFQPGGEPGEGPAGSGPMGPRGPPGKNGDDGEAGK 238

QY 58 GRPGEAGLP---GAKGLTSGSPGPP-----DGKTGPPGAGQDGRPPGPPGAR 105  
DB 239 GRPGERGPPGQARGLPSTAGLPGMKHRCFSLDGAAGDAGAGPKGPGSGENGAP 298

QY 106 GQAGVMGPPGKGAAGEPKAGER-----GVPPGPAVAGPA-----GKDGEAGAQ 150  
DB 299 GQMGRGLPGRGPRGAPGAPGARGNDGATGAAGPPGPTGPAAGPPGPPGAGKAGEAGPQ 358

QY 151 GPPGAPGAGRGPGP-----AGSPGFGGLGPA 180  
DB 359 GPRSEGGVGRGPPGPPGAPAGPAGNPDQCPKAGANGAPGIAGAPGPFAGRPS 418

QY 181 GPPGAGKPGQGVPGDLGAP-----GPSGPA-----GPFPGPT 213  
DB 419 GPQPGGPPGPKGNSGEPGAPSGKDTCAKEPGVGVGPPGPGAGEGKRGARBEPT 478

QY 214 GLPGRGPGGSGRPGADGAGPKGAPGERSPGAPGKSPGAPGAGPLGAK 273  
DB 479 GLPGRGPGGSGRPGADGAGPKGAPGERSPGAPGKSPGAPGAGPLGAK 538

QY 274 GLTSGSPGPDGKTGTPGAGQDGRFPFPGARGQAGVMGFPKGAAGFPKAGER 333  
DB 539 GLTSGSPGPDGKTGTPGAGQDGRFPFPGARGQAGVMGFPKGAAGFPKAGER 598

QY 334 GVRPAGVAGKVDGAGAGAGGPPGAPGAGERGEGQAGSPGQGLPAGPPGAGKP 393  
DB 599 GVPGPAGVAGKVDGAGAGAGGPPGAPGAGERGEGQAGSPGQGLPAGPPGAGKP 658

QY 394 GEQGVPGDLGAPGSGPAGE-----PGP-----TGLPGP 423  
DB 659 GEQGVPGDLGAPGSGPAGE-----PGP-----TGLPGP 718

QY 424 GERGGPSRPFPGADGAGVAGPKGAPGERSGPPGAPKSGPGE-----AGRPGAGLP 474  
DB 719 GSQAGPLQGMGPERGAAGLPKGRDAGAPKADGSPGKGVRLGTPIGPPGAPAG 778

QY 475 GAK-----GLTSGSPGPDGKTGTPGAGQDGRFPFPGARGQAGVMGFPKGAAGFPKAGER 515  
DB 779 GDKESGSPGAGTGGARGAGGEPGPPGAPGADGQPKAGEPDGAKGDA 838

QY 516 -----AGVMGPPGKGAAGFPKAGERGVPGPAA-----VGPAGKGEAGAGQPP 561  
DB 839 GPPGAPGAGPPGPIGVNAGAPGAKGARGSGAPGATGFPGAAGRVGPPGSGNAGPPGP 898

QY 562 GPAGP---AGERGQAGSPGQGLPAGPAGGEGAKGEGQVPGDLGAPGSPGAPGP 618

DB 899 GPAGKGGKGRGETGTPAGRPGEVGPFPAGKSGPAGKSGPAGADGAPAGCTGPPQGIAGOR 958  
QY 619 GPTGLPGRGSGRPFPGADGAGVAGPKGPPAGRGSPGAGP---KSGPGEAGRPGEA 675  
DB 959 GVVGLPQGRGERGFPGLPGRGSGRPFPGKQSGASGERGPPGMPGLAGPPGSGREGAP 1018  
QY 676 GLPFAKGLTSGSPGPDGKTGTPGAGQDGRFPFPGARGQAGVMGFPKGAAGEP 735  
DB 1019 GABGSPERGSGPAGKDRGETGTPAGPAGPAGPVPVGPAGKSGDRGETGTPAGPAGV 1078

QY 736 GKAGERGVFPFGAVGAPGAKDGE-----AGAQGPPGP-----AGPA 771  
DB 1079 GPAGARGPAGPQGRDKETGCEQDGRGKGRHGFSLGQGGPPGSGPGEQSPGASGPA 1138

QY 772 GERGEGPAGSP---GFQGLPAGPAGPAGKPGEGQVPGDLGAPGSPGAP 820  
DB 1139 GPRGPPGSAGAPKDKLNLGPIGPPGRGRTGADGVPVGPFPGPFP 1190

## RESULT 13

ABG93947  
ID ABG93947 standard; protein; 1461 AA.  
XX AC ABG93947;  
XX DT 26-NOV-2002 (first entry)  
XX DE Human polypeptide orthologous to DACC-11.  
XX KW Human; deer; rat; mouse; DACC; deer antler cartilage cell;  
KW cell stimulation; cell inhibition; cell growth; cell division;  
KW mesenchymal cell; chondrocyte; chondrogenesis; osteogenesis; growth;  
KW repair; regeneration; restoration; extracellular matrix;  
KW cartilaginous matrix; cartilage; disc; connective tissue; agonist;  
KW antagonist; gene therapy.  
XX OS Homo sapiens.  
XX PN WO200264625-A1.  
XX PD 22-AUG-2002.  
XX PF 15-FEB-2002; 2002WO-AU000163.  
XX PR 15-FEB-2001; 2001AU-00003116.  
XX PA (ADPP-) ADP PHARM PTY LTD.  
XX PA (UNSY) UNIV SYDNEY.  
XX PI Roubin R, Ghosh P;  
XX WIPI; 2002-643456/69.  
XX PT Stimulating or inhibiting cell growth and/or division, useful for  
PT stimulating chondrogenesis, cartilage, disc or connective tissue growth,  
PT repair, and/or regeneration, comprises administering deer antler  
PT cartilage gene.  
XX PS Claim 13; Page 154-160; 214pp; English.

The invention discloses a method for stimulating or inhibiting cell growth and/or division which comprises contacting or inserting into an animal cell a polypeptide comprising one of the deer antler cartilage cell (DACC) clones disclosed. More particularly, the method relates to these polypeptides stimulating mesenchymal cell growth and/or division and to transfecting these cells and chondrocytes with vectors carrying the genes of these polypeptides capable of stimulating chondrogenesis, osteogenesis, growth, repair, regeneration and/or restoration of the extracellular matrix. The chondrocytes selectively express genes required to form a cartilaginous matrix. The DACC polypeptides and polynucleotides are useful for identifying an agent that modulates the activity of the polypeptide, for stimulating mesenchymal cell growth and/or division by exposing animal mesenchymal cells to conditioned media or its active

CC fraction, obtained from deer antler cartilage cells, for inhibiting cell  
 CC growth and/or division by inserting into an animal cell, a compound which  
 CC inhibits the translation of the polynucleotide encoding the DACC. The  
 CC method and the polypeptides are useful for stimulating mesenchymal cell  
 CC growth and/or division or for stimulating chondrogenesis, cartilage, disc  
 CC or connective tissue growth, repair, regeneration and/or restoration in  
 CC an animal. The polynucleotides, polypeptides, agonists and antagonists  
 CC may be used in treatment modalities, specifically in gene therapy. The  
 CC polypeptides can be used as bait proteins in a two- or three-hybrid assay  
 CC to identify other proteins, which bind to or interact with the  
 CC polypeptide and are involved in modulating cell growth and/or division.  
 CC The sequences presented in ABG93923-ABG93948 are the proteins encoded by  
 CC the DACC cDNA clones  
 XX  
 XX Sequence 1461 AA;

Query Match 59.5%; Score 2761; DB 5; Length 1461;  
 Best Local Similarity 55.3%; Pred. No. 2.6e-162;  
 Matches 560; Conservative 31; Mismatches 229; Indels 192; Gaps 19;

QY 1 GPPGEGTGLPGRGGRGGRGFRGAGDGVAGPKPAGERSPPGAGKSPGEA--- 57  
 DB 176 GPMGSPGRGLPGPPGAGPGQGFQGPGEPCGECASQPMGPRGPPGKXNGDDGAGKP 235

QY 58 GRPCEAGLP---GAKGLTSPGSGP-----DKTGPAGQDGRPPGPPGAG 105  
 DB 236 GRPGERPPGQAGRLPGTAGLDMKCHRGFSGLDAKADGAPGKGFSPGENGAP 295

QY 106 GQAGVMFPFPGKAAGFPKAGER-----GVPPPGAVGPA-----GKDGAGAG 150  
 DB 296 GQMGPRGLPGRGRGAPGAPGAGARGNDGATCAAGPPGTPGAPPPGPAVAGKAGAPQ 355

QY 151 GPPGAPGAGERGQGP-----AGSPGFGQLGPA 180  
 DB 356 GPRGSEGVGVRGPPGPPGAGAGPAGNPGADQCPGAKGANGAPGAPGPPGARGPS 415

QY 181 GPPGAGKPGQGVFGDLGAP-----GFSGPA-----GEPGPT 213  
 DB 416 GPQFGPPGPKNSGFBGAPGSKDTCAGKEPGVGVQGPFGAGEGKRGARGEPPT 475

QY 214 GLPGRGGRGGRGFRGAGDGVAGPKPAGERSPPGAGKSPGAGRPGENGLPGAK 273  
 DB 476 GLPGRGGRGGRGFRGAGDGVAGPKPAGERSPPGAGKSPGAGRPGENGLPGAK 535

QY 274 GLTSGSPGPDGKTGTPGPPAGQDGRPPGPPGARGQAGVMGPPGKGAAGEPGKAGER 333  
 DB 536 GLTSGSPGPDGKTGTPGPPAGQDGRPPGPPGARGQAGVMGPPGKGAAGEPGKAGER 595

QY 334 GVPGPAGVAGPKDGEAGAGQCPGPPAGPAGERGEGQGPAGSPGQGLPGPAGPPGEAGKP 393  
 DB 596 GVPGPAGVAGPKDGEAGAGQCPGPPAGPAGERGEGQGPAGSPGQGLPGPAGPPGEAGKP 655

QY 394 GEQGVPGDLGAPGSPGAGE-----RGP-----TGLPGPP 423  
 DB 656 GEQGVPGDLGAPGSPGAGERGFRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 715

QY 424 GERGGGSRGFRGAGDGVAGPKPAGERSPPGAPKSPGE-----AGRPGEGALP 474  
 DB 716 GSQGFAPGLQMGFRGAGAGLPKPKDGRDAGPKGADGSPGKDGVRGLTGP:GPPGPAGAP 775

QY 475 GAK-----GLTSGSPGPDGKTGTPGPPAGQDGRPPGPPGARGQ----- 515  
 DB 776 GDKGESGSPGAPGTGARGAPDGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 835

QY 516 -----AGVMGPPGPKGAAGEPKKAGERGVGPPGPA-----VGPAGKDGAGAGQPP 561  
 DB 836 GPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 895

QY 562 GPAGP---AGERGEQGPAGSPGQGLPGPAGPPGAGKPGRQGVPGDILGAPGSPGAGEP 618  
 DB 896 GPAGKEGKGPRGTGPAGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 955

QY 619 GTGLPGRGGRGGRGFRGAGDGVAGPKPAGERSPPGAPG---KGSPEGACRPGEA 675

DB 956 GVVLPGQGRGFRGFRGLPGSPGEGKQKQPSGASGERGPPGPMGPPGLAGPESGREGAP 1015  
 QY 676 GLPGAKGLTSPGSPGPDGKTPGPPAGQDGRPPGPPGARGQAGVMGFFGPKGAGEP 735  
 DB 1016 GAEGSPGRDGSFGAKGDRGETGAPGAPGPGFVGPAKSGDRGETGTPAGPAGPV 1075

QY 736 KKAAGERGVGPPGAVGAPKADGE-----AGAQGPPGP-----AGPA 771  
 DB 1076 GVPARGPAGPQGRGKGTGEGQDRGKIKHGRFSGLGQGPFGPPGSPGEGSPGASGPA 1135

QY 772 GERGQGPAGSP---GFQGLFPGAPGEGEAGKPGEQGVGPDLAGAPSGSPAG 820  
 DB 1136 GPRGPPGSAGAPGKDGGLGFLPGIPGPPGRTGDPAGVGPVGPFPFGPPGPPG 1187

## RESULT 14

AAU14136

XX AAU14136 standard; protein; 1464 AA.

AC AAU14136;

XX DT 24-OCT-2001 (first entry)

XX DE Human novel protein #7.

XX KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;  
 KW immunomodulatory; cycostatic; neuroprotective; vulnerary; nootropic;  
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;  
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;  
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
 KW tissue regeneration; immune disorder.

XX OS Homo sapiens.

XX PN WO200155437-A2.

XX PD 02-AUG-2001.

XX PF 25-JAN-2001; 2001WO-US002623.

XX PR 25-JAN-2000; 2000US-00491404.

XX PA (HYSE-) HYSE INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-451939/48.

XX PT N-PSDB; AAS22441.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,  
 PT nervous system disorders, and for regenerating bone and cartilage.

XX PS Example 4; Page 525-527; 894pp; English.

XX CC The invention relates to polynucleotides encoding novel human proteins or  
 CC their active domains. The polypeptides, polynucleotides and antibodies  
 CC raised against the polypeptides are used in a method of treatment of a  
 CC mammal and prevention of disorders caused by the aberrant protein  
 CC expression or activity. The polypeptides can be used as molecular weight  
 CC markers, food supplements, and in antibody production. The polypeptides  
 CC are used to identify compounds which bind to the polypeptides.  
 CC Polynucleotides of the invention are used as probes and primers, for  
 CC sequencing, for chromosome or gene mapping, in the production of  
 CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene  
 CC therapy. Polypeptides of the invention can be used to target drugs to a  
 CC tumour, in assays to determine biological activity, to raise  
 CC antibodies/ elicit an immune response, to determine quantitative protein  
 CC levels, as tissue markers, and to isolate receptors or ligands.  
 CC Polypeptides of the invention may also be useful in treating platelet  
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting

CC the proliferation, differentiation and survival of stem cells, as a  
 CC contraceptive, treating osteoporosis and osteoarthritis, aneemia,  
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-  
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory  
 CC diseases, nervous system disorders, and infection. The present sequence  
 CC represents a protein of the invention  
 XX  
 XX Sequence 1464 AA;

Query Match 59.5%; Score 2761; DB 4; Length 1464;  
 Best Local Similarity 55.3%; Pred. No. 2.6e-162; Indels 192; Gaps 19;  
 Matches 560; Conservative 31; Mismatches 229;  
 QY 1 GPPGPGTGLPFPGERGSGRFPFGADGVAGPKGAPGAGSGPPGAGPKGSGPGEA--- 57  
 DB 179 GPMGSGPRGLPFPAGPQGFQGPCEGPGASGPMGPRGPPGPKNGDDGEAGKP 238  
 QY 58 GPGGAGLP---GAKLTGSGSPCP-----DGTGPPGAGQDGRPPGPPGAR 105  
 DB 239 GRPGRGPPGPGARGLPGLTGLPKMGRHFGSLDCAKGDAGPAGPKGPPSGENGAP 298  
 QY 106 GQAGVMGFPKGAAGBEGKAGER-----GVPPPGAVGPA-----GKDGAGAG 150  
 DB 299 QMGPRGLPGRGRFGAPGAGAGNDGATCAAGPFGPTGPPGPPGAVGAGKEAGPQ 358  
 QY 151 GPPGAPGAGRGEBOGP-----AGSPFGQLPFGPA 180  
 DB 359 GPRSEGGQVGRGFPFGPAGAGPAGNPGADQCPGAKGANGAPGIAGAPFPFGARGPS 418  
 QY 181 GPGGAGKPGQGVGDLGAP-----GPGSGPA-----GERGPT 213  
 DB 419 GPQGGGPPGPKNGSGEPGAPGSGKGTGAKGEPDVPVQGGPPGABEGKRGARGSPGPT 478  
 QY 214 GLPDPGERGPGSRGFPAGDVAGPKGAPAGERSGPPAGPKGSPGAGRPGAGLPGAK 273  
 DB 479 GLPDPGERGPGSRGFPAGDVAGPKGAPAGERSGPPAGPKGSPGAGRPGAGLPGAK 538  
 QY 274 GLTSGSGPDGKTGTPGAGDGRGPPGPPGARGQAGVMGPPGPKGAAGPPGKAGER 333  
 DB 539 GLTSGSGPDGKTGTPGAGDGRGPPGPPGARGQAGVMGPPGPKGAAGPPGKAGER 598  
 QY 334 GVPFPAGVAGKDXGAGAGPAGPAGAGERQAGPAGSPGQGLPAGPAGPAGKRP 393  
 DB 599 GVPFPAGVAGKDXGAGAGPAGPAGAGERQAGPAGSPGQGLPAGPAGPAGKRP 658  
 QY 394 GEQVPGDLGNPSPGAGE-----PGP-----TGLPGPP 423  
 DB 659 GEQVPGDLGNPSPGARGRPFGERGVQGPFPAGPRGANGAPNGDCAKGDAGAPGAP 718  
 QY 424 GERGGPSRFPFGADGVAGPKGAPAGERSGPPAGPKGSPGAGRPGAGLPGAP 474  
 DB 719 GSQAGPLQGWGREGAAGLPKGDGADGPKGADGSPKDGVRGLTGPFGPAGAP 778  
 QY 475 GAK-----GLTSGSGPDGKTGTPGAGDGRGPPGPPGARGQ----- 515  
 DB 779 GDKGESGFPAGPTGARGAPGDRGEPGPPGAGFAGPAGDQCPGAKGEPGDAGAKDA 838  
 QY 516 ----AGVMGFPKGAAGPPGKAGERGVPPGPA-----VGPAGKDXGAGQPP 561  
 DB 839 GPPGAPGAPGPPGIGNVGAPGAKGAGSAGPPGATGFGCAAGRVGPPGSGNAGPPGP 898  
 QY 562 GPAGP---AGERQAGPAGSPGQGLPAGPAGPAGKPGEQQVPGDGLGAPGSPGAPGP 618  
 DB 899 GPAGKEGKGRGTGTPGAGRPGVPPGPPGAGKSGPGADGAPAGPTGPPQGIAGOR 958  
 QY 619 GPTGLPFPGERGSGRFPAGDVAGPKGAPAGERSGPPAGP---KSPGAGRPGPA 675  
 DB 959 GVVGLPQQRGERGFPGLPSPSGRFGKQSGSGRGGPPGPPGPPGLAGPPGESREGAP 1018  
 QY 676 GLPAGKLTGSPSGPDGKTGTPGAGDGRGPPGPPGARGQAGVMGFPKGAAGBP 735

DB 1019 GAEGSPRDSGPGAKGDRGTGTPAGPPGAPGAPGPPVPGAGKSGDRGTGTPAGPVP 1078  
 QY 736 GKAGRGVPGPCAVCPACKDGE-----AGAGGPPGP-----AGPA 771  
 DB 1079 GPVAGRPPAGPQPRDKGTGEQDGRGKGRHSGSLQGPPGPGSGEQPGSGASGPA 1138  
 QY 772 GERGEQGPAGSP---GFOGLPGPAGPPGAGKPGGQGVPGDLCAPGSPGAP 820  
 DB 1139 GPRGPPGSAGAPGKDLNGLPGPIGPPGPRGTGAGPVGPPGPPGPPGPPG 1190

## RESULT 15

AAV84541

ID AAY84541 standard; protein; 1057 AA.

XX AAY84541;

XX AC (first entry)

XX DT 25-JUL-2000

XX DE Amino acid sequence of a human collagen 1 (alpha1) protein.

XX KW Extracellular matrix protein; self aggregation; hydroxylated proline;

XX KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;

XX KW collagen; fibrinogen; fibronectin; post translational hydroxylation.

XX OS Homo sapiens.

XX PN EP992586-A2.

XX PD 12-APR-2000.

XX PP 07-OCT-1999; 99EP-00119184.

XX PR 09-OCT-1998; 98US-00169768.

XX PA (USUS) US SURGICAL CORP.

XX PI Gruskin EA, Buechter DD, Zhang G, Connolly K;

XX DR WPI; 2000-259138/23.

XX DR N-PSDB; AAA12502.

XX PT Production of extracellular matrix proteins containing 4-trans-

XX PT hydroxyproline results in native self aggregating proteins, useful on

XX PS medical implants.

XX PS Disclosure; Fig 27A-E; 260pp; English.

XX CC The specification describes a method for producing an extracellular  
 CC matrix protein or its fragment. The extracellular matrix protein is  
 CC capable of self aggregating in a cell which does not ordinarily  
 CC hydroxylated prolines. The method comprises optimising a nucleic acid  
 CC sequence for expression in the cell by substitution of codons preferred  
 CC by that cell for naturally occurring codons not preferred by the cell;  
 CC incorporating the nucleic acid sequence into the cell; and contacting the  
 CC cell with a hypertonic growth medium containing at least one amino acid,  
 CC selected from the group consisting of trans-4-hydroxyproline and 3-  
 CC hydroxyproline to allow at least one of the amino acids to be assimilated  
 CC into the cell and incorporated into the extracellular matrix protein. The  
 CC method may be used to make host cells assimilate and incorporate trans-4-  
 CC hydroxyproline into proteins. This is especially useful in the  
 CC recombinant production of proteins such as collagen, fibrinogen and  
 CC fibronectin whose ability to self aggregate and produce functional  
 CC proteins depends on the post translational hydroxylation of proline. The  
 CC method is also useful in studying the structure and function of  
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The  
 CC present sequence represents a human collagen 1 (alpha1) protein, which  
 CC may be produced using the method of the invention

XX SQ Sequence 1057 AA;

Query Match 59.5%; Score 2759; DB 3; Length 1057;  
 Best Local Similarity 55.3%; Pred. No. 2.7e-162;

Matches		560;	Conservative	31;	Mismatches	229;	Indels	192;	Gaps	19;
Qy	1	GPPGEPPTGLPDPGGRGPGSRGPPGADGVAGPKPAGERGSGSPGAPKSPGCEA--- 57								
Db	18	GPMGSPGSRGLPDPGPPGAPGQGGPPGEPGEGFASGPMGPRPPGPPGKXGDDGAGKP 77								
Qy	58	GRPEAGLP---GAKGLTSPGSGFP-----DGKTGPPGACQDGRPPGPPGAR 105								
Db	78	GRGERGPPGQARGLPCTAGLPKMKHHRGFSGLDKAKGDAGFAGPKGEPGSPGENGAP 137								
Qy	106	GQAGVMGFPKGAAGEFGKAGER-----GVFGPPGAVGPA-----GKDGAGAQ 150								
Db	138	GQMGRGLPGERGRPGAPGAGAGNDGATGAAGPPGPTGPGAPPPGAVGAKGAGGPO 197								
Qy	151	GPPGAPGAGERGEQF-----AGSPGFQGLPGPA 180								
Db	198	GPRSEGFQGVGEGPPGPPAGAAAGPACNPDGQPCAKGANGAPGTAGAPGPPGARGPS 257								
Qy	181	GPPGAGKPGEGQVPGDLGAP-----GPSGPA-----GPPGPT 213								
Db	258	GPQGGGPPGPKGNSGEPGAFSGKDTGAKGEGPGVGVQPPGPAEGEKGARGEPGPT 317								
Qy	214	GLPDPGERGPGSRGPGADGVAGPKPAGERGSGPAGPKGSPGAGRPGEAGLPKAK 273								
Db	318	GLPDPGERGPGSRGPGADGVAGPKPAGERGSGPAGPKGSPGAGRPGEAGLPKAK 377								
Qy	274	GLTSPGSGPDGKTGPPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGFPKAGER 333								
Db	378	GLTSPGSGPDGKTGPPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGFPKAGER 437								
Qy	334	GVPDPGAVGAGKDGAGAGQPPGPPGAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKP 393								
Db	438	GVPDPGAVGAGKDGAGAGQPPGPPGAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKP 497								
Qy	394	GEQGVGDLGAPGSPGAGE-----PGP-----TGLFPPP 423								
Db	498	GEQGVGDLGAPGSPGAGERGFPGERGVQPPGPPGARGANGAPGNDGAKGDAGAPGAP 557								
Qy	424	GERGPGSRGPPGADGVAGPKPAGERGSGPAGPKGSPGE-----AGRPGEAGLP 474								
Db	558	GSQAGPGLQGMFGERGAAGLPKPKGDRDAGKADGSGPKDGVRLTGTPTGPPGAPGAP 617								
Qy	475	GAK-----GLTSGSGSPGDKTGPAGQDGRPPGPPGARGQ----- 515								
Db	618	GDKSGSGSPGAPGTGARGAPGDRGEPGPPGPPGAPGPPGADGQPGKAGEPGDAGAKGDA 677								
Qy	516	-----AGVMGPPKGAAGEPKKAGERGVPPPPGA-----VGPAKDKGEAGAGPP 561								
Db	678	GPPGAPGAPGPPGPIGNVAGPCKAKGARGSAGPPGATGPPGAAGRVGPPGSGNAGPPGPP 737								
Qy	562	GPAGP-----AGERGEQGPAGSPGQGLPGPAGPPGEPAGKPEGQGVPCDLGAPGSPGAGEP 618								
Db	738	GPAGKEGKGPRGETGPAGRPGEVGPPGPPGAGEKSGFADGPPAGAPGTGPGQIAGQR 797								
Qy	619	GPTGLPFPGERGPGSRGPPGADGVAGPKPAGERGSGPPAGP---KGSFGEAGRPGEA 675								
Db	798	GVVGLPQGRGERGFPGLPSPGEPKQGPSGASGERGPPGMPGPPGLAGPPGSGREGAP 857								
Qy	676	GLPGAAGLTPGSPGPDGKTGPAGQDGRPPGPPGARGQAGVMGFPKPKGAAGEP 735								
Db	858	ALEGSPDRDGSFCAKGDGRTGETGAPGPPGAPGAPGVPVGPAGKSGDRGETGPPAGPAPV 917								
Qy	736	GKAGERGVPPGCAVCPACKDGE-----AGAGGPPGP-----AGPA 771								
Db	918	GPAGARGPAGPQGRGDKGETGEQDGRGKHKHRSGLQGLPPGSPGEGQPSGASGPA 977								
Qy	772	GERGEQGPAGSP---GFQGLPGPAGPPGAGKPGEQGVGDLGAPGSPGAPG 820								
Db	978	GFRGPPGSAGAPKDKGLNGLPGPIGPPGPRGTGADGAVGPPGPPGPPGPPG 1029								

Search completed: October 18, 2004, 13:37:32  
Job time : 135.602 secs

*This Page Blank (uspio)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:31:20 ; Search time 32.5431 Seconds  
(without alignments)  
1673.075 Million cell updates/sec

Title: US-10-658-989A-3  
Perfect score: 4640  
Sequence: 1 GPGGPGPTGLPDPGRRG.....GEOGVFDLGAQPSGAG 821

Scoring table: ELOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2765	59.6	1464	4	US-09-331-347C-21
2	2752	59.3	1461	4	US-09-585-887-9
3	2752	59.3	1461	4	US-09-289-578-9
4	2746	59.2	1057	3	US-08-931-820-1
5	2704	58.3	1341	3	US-08-963-825-18
6	2704	58.3	1341	3	US-08-500-811-18
7	2704	58.3	1341	3	US-09-570-573-18
8	2704	58.3	1341	3	US-08-548-608-18
9	2511	54.1	822	3	US-09-219-849-49
10	2488	53.6	1017	4	US-08-468-996-10
11	2478	53.4	1064	1	US-08-642-255-62
12	2476	53.4	1060	3	US-08-931-820-3
13	2476	53.4	1418	3	US-08-963-825-20
14	2476	53.4	1418	3	US-09-010-999-1
15	2476	53.4	1418	3	US-08-500-811-20
16	2476	53.4	1418	3	US-09-570-573-20
17	2476	53.4	1418	3	US-08-548-608-20
18	2464	53.1	1442	2	US-08-316-650-12
19	2464	53.1	1442	5	PCT-US95-02251-12
20	2370	51.1	1057	3	US-08-931-820-4
21	2350.5	50.7	1078	3	US-08-963-825-21
22	2350.5	50.7	1078	3	US-08-500-811-21
23	2350.5	50.7	1078	3	US-09-570-573-21
24	2350.5	50.7	1078	3	US-08-548-608-21
25	2267.5	48.9	1366	3	US-08-963-825-19
26	2267.5	48.9	1366	3	US-09-500-811-19
27	2267.5	48.9	1366	3	US-09-570-573-19

28 2267.5 48.9 1366 3 US-09-548-608-19  
29 2265.5 48.8 1366 4 US-09-585-887-10  
30 2265.5 48.8 1366 4 US-09-289-578-10  
31 2255.5 48.6 1024 3 US-08-931-820-2  
32 2254.5 48.6 1065 1 US-08-642-255-72  
33 2169 48.7 1806 4 US-09-919-497-56  
34 2156.5 46.5 960 3 US-09-219-849-5  
35 2149.5 46.3 777 1 US-08-642-255-53  
36 2144.5 46.2 720 3 US-09-219-849-4  
37 2123.5 45.8 897 1 US-08-397-633A-50  
38 2096 45.2 595 3 US-09-219-849-48  
39 2096 45.2 595 3 US-09-219-849-50  
40 2057 44.3 837 1 US-09-795-061-4  
41 2055 44.3 837 1 US-08-175-155-68  
42 2055 44.3 837 1 US-08-477-509B-103  
43 2055 44.3 837 1 US-08-642-255-101  
44 2055 44.3 837 2 US-08-707-237A-75  
45 2055 44.3 837 3 US-08-482-085B-103

#### ALIGNMENTS

##### RESULT 1

US-09-331-347C-21

; Sequence 21, Application US/09331347C

; Patent No. 6617431

; GENERAL INFORMATION:

; APPLICANT: Meristem Therapeutics, S.A.

; APPLICANT: Meristem Therapeutics, S.A.

; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, and

; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, and

; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, and

; FILE REFERENCE: 1149-3

; CURRENT APPLICATION NUMBER: US/09/331,347C

; CURRENT FILING DATE: 1999-08-17

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 21

; LENGTH: 1464

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-09-331-347C-21

Query Match 59.6%; Score 2765; DB 4; Length 1464;

Best Local Similarity 55.4%; Pred. No. 2.9e-171;

Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPGGPGPTGLPDPGRRGPPGADGVAGPKPAGERGSPGAPKSPGGEA--- 57  
DB 179 GPMGSGPRGLPDPGAPGPGFQGPGEPPGSGPMGPRGPPGPKNGDDGEAGKP 238  
QY 58 GRPGAGLP---GAKGLTSGSPGSP---DGKTGPPGPAQDGRPPGPPGAR 105  
DB 239 GRPGERGPPGQARGCLPOTAGLPGMKHGRGSLDGAAGDAGPAGKPGSPGENGAP 298  
QY 106 GQGVGVGPPGPKAAGEPKAGER-----GVPPGPAVGPA-----GKDGEAGNQ 150  
DB 299 GQMGPRGLPGERGPPGAPGAGNAGTGAAGPFGTGTGTPGPPGPAVAKGAGEGQ 358  
QY 151 GPGGPAAPAGERGEOGP-----AGSPGFQGLPGPA 180  
DB 359 GRFGSEPGVGVGPPGPPGAPAGAGNPGADGPGAKGANGAGIAGAPFPFGARGPS 418  
QY 181 GPPGEAGKPGEOGVGPDLGAP-----GPSGA-----GEPGPT 213  
DB 419 GPGGPGPPGPKNGSGEPGAGSKDGTAKGPPGVGVQGPFGPAGEGKRGARGEPGPT 478  
QY 214 GLPGRGERGSGSPGPPGADGVAGPKPAGERGSPGAPKSGPCEAGRPGCEAGLPKAK 273  
DB 479 GLPGRGERGSGSPGPPGADGVAGPKPAGERGSPGAPKSGPCEAGRPGCEAGLPKAK 538  
QY 274 GLTSGSPGPPGPKTTPPGAPQDGRPPGPPGPPGAGQAGVNGFPFGPKGAAGEPGKAGRR 333

Db 539 GLTSGSPGPDGKTTGPPGPAQDGRPPGPPGARGQAQVMGPPGPKGAAGFPGKAGER 598  
Qy 334 GVPPPGCAVAGPAGKDGAGAGCPGPPAGPAGEREQGPPGPGFQGLPGPAGPPGAGKP 393  
Db 599 GVPPPGCAVAGPAGKDGAGAGCPGPPAGPAGEREQGPPGPGFQGLPGPAGPPGAGKP 658  
Qy 394 GEQVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423  
Db 659 GEQVPGDLGAPGSPGAGEGFFGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 718  
Qy 424 GERGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 474  
Db 719 GSQAPGLQMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 778  
Qy 475 GAK-----GLTSPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 515  
Db 779 GDKGESGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 838  
Qy 516 -----AGVMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 561  
Db 839 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 898  
Qy 562 GPAGP-----AGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 618  
Db 899 GPAGKEGKPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 958  
Qy 619 GPTGLPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 675  
Db 959 GVVGLPQGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1018  
Qy 676 GLPAGKGLTSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 735  
Db 1019 GAEGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1078  
Qy 736 KGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 771  
Db 1079 GPAGARGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1138  
Qy 772 GERGEQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 820  
Db 1139 GPRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1190

## RESULT 2

US-09-585-887-9  
; Sequence 9, Application US/09585887  
; Patent No. 6413742  
; GENERAL INFORMATION:  
; APPLICANT: Olsen, David R  
; APPLICANT: Chang, Robert  
; APPLICANT: McMullin, Hugh  
; APPLICANT: Hitzeman, Ronald A.  
; APPLICANT: Chisholm, George  
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND  
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT  
; TITLE OF INVENTION: CELLS  
; FILE REFERENCE: 225002030400  
; CURRENT APPLICATION NUMBER: US/09/585,887  
; CURRENT FILING DATE: 2000-05-31  
; PRIOR FILING DATE: 09/289,578  
; PRIOR FILING DATE: 1999-04-09  
; PRIOR APPLICATION NUMBER: 60/084,828  
; PRIOR FILING DATE: 1998-05-08  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 1461  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-585-887-9

Query Match 59.3%; Score 2752; DB 4; Length 1461;  
Best Local Similarity 55.2%; Pred. No. 2e-170;

Matches 559; Conservative 31; Mismatches 230; Indels 192; Gaps 19;  
Qy 1 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 57  
Db 176 GPMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 235  
Qy 58 GRPGEAGLP--GAKGLTSGSPGSPG-----DKTTPGPPGPPGPPGPPGPPGPPGPP 105  
Db 236 GRPGEAGLP--GAKGLTSGSPGSPG-----DKTTPGPPGPPGPPGPPGPPGPPGPP 295  
Qy 106 GQAGVMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 150  
Db 296 GQAGVMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 355  
Qy 151 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 180  
Db 356 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 415  
Qy 181 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 213  
Db 416 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 475  
Qy 214 GLPFPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 273  
Db 476 GLPFPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 535  
Qy 274 GLTSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 333  
Db 536 GLTSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 595  
Qy 334 GVPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 393  
Db 596 GVPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 655  
Qy 394 GEQVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423  
Db 656 GEQVPGDLGAPGSPGAGEGFFGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 715  
Qy 424 GERGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 474  
Db 716 GSQAPGLQMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 775  
Qy 475 GAK-----GLTSPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 515  
Db 776 GDKGESGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 835  
Qy 516 -----AGVMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 561  
Db 836 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 895  
Qy 562 GPAGP---AGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 618  
Db 896 GPAGKEGKPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 955  
Qy 619 GPTGLPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 675  
Db 956 GVVGLPQGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1015  
Qy 676 GLPAGKGLTSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 735  
Db 1016 GAEGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1075  
Qy 736 KGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 771  
Db 1076 GPVARGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1135  
Qy 772 GERGEQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 820  
Db 1136 GPRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1187

RESULT 3  
US-09-289-578-9

```

; Sequence 9, Application US/09289578
; Patent No. 6428978
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/289,578
; PRIOR FILING DATE: 1999-04-10
; PRIOR APPLICATION NUMBER: 60/084,838
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-289-578-9

```

## RESULT 4

US-08-931-820-1  
; Sequence 1, Application US/08931820  
; Patent No. 6010863  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Assay for collagen degradation  
; NUMBER OF SEQUENCES: 4  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/931,820  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 96202596.1  
; FILING DATE:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1057 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; TISSUE TYPE: Collagen type I  
US-08-931-820-1

Query Match	59.2%	Score	2746;	DB 3;	Length	1057;			
Best Local Similarity	56.3%;	Pred. No.	3.7e-170;						
Matches	556;	Conservative	32;	Mismatches	232;	Indels	168;	Gaps	18;
Qy	1	GPPCEP	-----GFTGLPGPGRGGSGCFGADGVAGPKGPAGERGSP	-----	45				
Db	42	GPPCEP	CGASGMPGRGPPGPKNGDGEAKGKPGRGERGPPGQGARGLPCTAGLP	101					
Qy	46	-----	GPAGPK	---GSPGEAGRPGCEAGLPAGKAGTGTGSPGSPG	---	80			
Db	102	GMKGRGFGSLDGA	KADGAGPAGKPEFGSPCENGAPGQMGPRGLPGRGRPGGAPGAGAR	161					
Qy	81	--DKGT	--GPPGAGQDGRPPGPPPGAGQA	-----	GMWGFPCKGAGEP	123			
Db	162	GNDGATGAAGPPG	TGPAGPPGFGVCAKGAAGPQGRGSEGQGVRRGFP	221					



Db 601 GLQWPGERGAAGLPKPGDRGDAGPKGADGSPKGDVGRGLTGPIGPPGAGAPGDKGES 660  
Qy 478 -----GLTSPGSGPDGKTPGPPGAGQDGRPGPPPGARGQ-----A 516  
Db 661 GPSGPAFTGARGAPGDRGPPGPPGAGFAGPPGADGQPGAKGEPDGAGAKGAGPPGPA 720  
Qy 517 GVMGFPKGAAGPPGKAGERGVPGPFGA-----VGPAGKDGAGAGQGGPPGAGP- 566  
Db 721 GPAGFPPIGNVGAAGKAGARGAGFPATGCFGAAGRVGPPGPGNAGPPGPPGAGKE 780  
Qy 567 --AGERGQGPAGSGFQGLGPPGAPGGEAGKPGQGVGPDGLGAPGSGPAGSGFTGLP 624  
Db 781 GKGPRGETGAPRGVGVGPPGPPGAGKSGSPGADGAPAGTTPGQAGQGVVGLP 840  
Qy 625 GPPCERGSGSRGPPGADGAVAGKPGAGERSGPPAGP---KGSPPGAGRPGAGLPKAG 681  
Db 841 GQERGERFGLPGPSGPPGKQGSASGERGPPGMPGPPGLAGPPGSGREGAPGAGSP 900  
Qy 682 GLTSPGSGPDGKTPGPPGAGQDGRPGPPGAGQAGVGMGPPGKGAAGFPKAGER 741  
Db 901 GRDGSFGAKGDRGETGAPGPPGAGXGAXGAPGVGPPGAKSGDRGETGAPGAPGVGPPAGAR 960  
Qy 742 GVPGPPGAVGAPGAKDGE-----AGAQQGPPG-----AGPAGRGQ 777  
Db 961 GPAGPQPRGDKGTGEGQDRGIKGRGSGGLQGGPPGPPGSGQSGASGAPGPPG 1020  
Qy 778 GPAGSP---GPGGLPGAPGGEAGKTPGQGVGPDGLGAPGSPGAG 820  
Db 1021 GSAGAPGKGLNLGPIGPPGPRGRTCDAGVPVPPGPPGPPG 1066

RESULT 6  
US-09-500-811-18  
; Sequence 18, Application US/09500811  
; Patent No. 6323314  
; GENERAL INFORMATION:  
; APPLICANT: Quist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/500,811  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/187,319  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ggoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1341 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: COLLAGEN ALPHA 1 (I)  
; US-09-500-811-18  
  
Query Match 58.3%; Score 2704; DB 3; Length 1341;  
Best Local Similarity 57.3%; Pred. No. 2.3e-167;  
Matches 542; Conservative 28; Mismatches 244; Indels 132; Gaps 15;  
  
Qy 1 GPPGE---PGPTGLPGPPCERGGPPGSRGF-----PGADGVAGPKGPPGAGER 42  
Db 127 GPPGRDGLPGQGLPGPPGPPGPPGPPGLGGNFAPQLYCYDEKSTGGISVPGPMGSPGR 186  
Qy 43 GSPGAPGKSGPGEAGRPGCEAGLPKAGLGTSPGSGPDGKTPGPPGAGQDGRP---GPP 99  
Db 187 GLGPPGAPGPPZGFZGPPGZGPGASGPMGPPGPPGPPGKGBBGGZAGKPPGZRPP 246  
Qy 100 GPPGARGQAGVGMGPPGKGAAGEPKGAGERGVPGPVAVGVPAGKDGAGAGQPPGAPGA 159  
Db 247 GPZGARGLPCTAGLPQMKHRCFSLBGAKGBAGPAGPKGZPSGSGZSCAPGZMGPPGPK 306  
Qy 160 GERGGQGPAGSPGFOGLPGPAGPGEAGKPGSQGVGPDGLGAPGPPGSGPAGSPGFTGLP 219  
Db 307 GNSGE-----PCAPGSKGDTGAKGEPGVGVQGGPPGAGGEGKRGARGEPGPTGLP 360  
Qy 220 GERGGSGSRGFPAGDGVAGKPGAGERSGPPGAGKSGSPGAGRPGCEAGLPKAGLGTSP 279  
Db 361 GERGGSGSRGFPAGDGVAGKPGAGERSGPPGAGKSGSPGAGRPGCEAGLPKAGLGTSP 420  
Qy 280 GSPGPDGKTPGPPGAPQDGRPPGPPGARGQAGVGMGPPGPKGAAGEPKKAGERGVP 339  
Db 421 GSPGPDGKTPGPPGAPQDGRPPGPPGARGQAGVGMGPPGPKGAAGEPKKAGERGVP 480  
Qy 340 GAVGAPGKDGAGAGQAGPPGAPGAGERGEOGAGSPGQGLPGPAGPGEAGKPGEOGV 399  
Db 481 GAVGAPGKDGAGAGQAGPPGAPGAGERGEOGAGSPGQGLPGPAGPGEAGKPGEOGV 540  
Qy 400 GDLGAPGPPGPPAGE-----PGS-----TGLPPPGPPGPPG 429  
Db 541 GDLGAPGPPGPPAGE-----PGS-----TGLPPPGPPGPPG 600  
Qy 430 GSRGPPGADGVAGKPGAGERSGPPGAPKSGPGE-----AGRPGCEAGLPKAG--- 477  
Db 601 GLQWPGERGAAGLPKPGKDRDAGPKGADGSGKDGVRGLTGPIGPPGAGAPGDKGES 660  
Qy 478 -----GLTSPGSGPDGKTPGPPGAGQDGRPGPPGARGQ-----A 516  
Db 661 GPSGPAFTGARGAPGDRGEPGPPGPPGAGPDGQPGAKGEPDGAGAKGAGPPGPA 720  
Qy 517 GVMGFPKGAAGEPKAGERGVPGPFGA-----VGPAGKDGAGAGQGGPPGAGP- 566  
Db 721 GPAGFPPIGNVGAAGKAGARGAGPPGATGTFPGAAGRVGPPGPGNAGPPGPPGAGKE 780  
Qy 567 --AGERGQGPAGSGFQGLPGPAGPGEAGKPGQGVGPDGLGAPGSGPAGSGFTGLP 624  
Db 781 GKGPRGETGAPRGVGVGPPGPPGAGKSGSPGADGAPAGTTPGQAGQGVVGLP 840  
Qy 625 GPPCERGSGSRGFPGADGAVAGKPGAGERSGPPAGP---KGSPPGAGRPGAGLPKAG 681  
Db 841 GQERGERFGLPGPSGPPGKQGSASGERGPPGMPGPPGLAGPPGSGREGAPGAGSP 900  
Qy 682 GLTSPGSGPDGKTPGPPGAGQDGRPGPPGAGQAGVGMGPPGKGAAGFPKAGER 741  
Db 901 GRDGSFGAKGDRGETGAPGPPGAGXGAXGAPGVGPPGAKSGDRGETGAPGAPGVGPPAGAR 960  
Qy 742 GVPGPPGAVGAPGAKDGE-----AGAQQGPPG-----AGPAGRGQ 777  
Db 961 GPAGPQPRGDKGTGEGQDRGIKGRGSGGLQGGPPGPPGSGQSGASGAPGPPG 1020

```
QY 778 GPAGSP---GFGGLPGAPGPGGAGKPGGQGVGDLGAPGSPGAG 820
Dd 1021 GSAGAPGKGLNGLPFGPIGPGPRGRTGDAGVGGPPGPPGPPG 1066

RESULT 7
US-09-570-573-18
; Sequence 18, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESS: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
US-09-570-573-18

Query Match 58.3%; Score 2704; DB 3; Length 1341;
Best Local Similarity 57.3%; Pred. No. 2.3e-167;
Matches 542; Conservative 28; Mismatches 244; Indels 132; Gaps 15;

QY 1 GPPGE---PGPTGLPGPPGPGGPGSGRGF-----PGADGVAGPKPGAGER 42
Dd 127 GPPCRDGIQPGGLPGPPGPPGPPGLGNFAPQLYDEKSTGGISVPGWPGSGPR 186
QY 43 GSPGAPGKSPGAGRPAGLPGAKGLTSGSPGPGDKTGPAGAGDGRP---GPP 99
Dd 187 GLPGPPGAPGPGZGPGPGZPGZPGAGSPGMPGPRGPPGPKBGBZAGKPGRPGZRGPP 246
QY 100 GPPGARGOAGVWPGPPGKGAAGEGKAGRGVPPGAVGAPGAGKGAAGAGGAGGPPGAPGA 159
Dd 247 GPPZGARGLPGTAGLPGMKHGRGSLGKAGBAGPAGKGPZPGZBAGPZMGPPGPK 306

160 GERGEGGAGSPGFGGLPGAPGPGGAGKPGGQGVGDLGAPGSPGAGBPPTGLPGPP 219
Dd 307 GNSGE-----PGAPGSKGDTCAKGEPOFVGVQVQPPGPGAGEGKRGAGEGPTGLPGPP 360
QY 220 GERGGGSRGFPFGADGVAGPKXGPAGERGSPGAPGKSGPGEAGRPGEAGLGAAGLGTSP 279
Dd 361 GERGGGSRGFPFGADGVAGPKXGPAGERGSPGAPGKSGPGEAGRPGEAGLGAAGLGTSP 420
QY 280 GSPGPDGKTGPPGPGAGQDGRPGPPGARGOAGVWPGPPGKGAAGEGKAGRGVGP 339
Dd 421 GSPGPDGKTGPPGPGAGQDGRPGPPGARGOAGVWPGPPGKGAAGEGKAGRGVGP 480
QY 340 GAVGPAKDXDGEAGAGQPPGPAERGERGEOGPPGAPGPGFQGLPGAGPPGGEAGKPGEGGVP 399
Dd 481 GAVGPAKDXDGEAGAGQPPGPAERGERGEOGPPGAPGPGFQGLPGAGPPGGEAGKPGEGGVP 540
QY 400 GDLGAPGSGPAGE-----PGP-----TGLPGPPGERGPP 429
Dd 541 GDLGAPGSGPARGERGERGPPGPPGARGOAGVWPGPPGANGAPGNDGAKGADGAPGAPGSOAP 600
QY 430 GSRGPPGADGVAGPKGPAERGSPGAPGKSGPGE-----AGRPGEAGLPGAK--- 477
Dd 601 GLQGMFGERGAAGLPGPKGDRGADGPKGDSGPKDGVRLTGTGTPGPPGAGAGDKGES 660
QY 478 -----GLTSPGSPGPDGKTGPPGAGQDGRPGPPGARGO-----A 516
Dd 661 GPPSGPAGTARGAPGDRGEPGPPGAPGAGPPGADGQPGAKGPPGDAGAKGDAGPPGPA 720
QY 517 GVMGPPGPKGAAGPKGAGRGVGP 566
Dd 721 GPAGPPGPIGNVAPGAKGARGSPGPPGATGFPGAAGRVGPPGSGNAGPPGPPGPAKE 780
QY 567 --AGRGEGGAGSPGFGFQGLPGPAGPGEAGKPGEQGVGDLGAPGSPGAGBPPTGLP 624
Dd 781 GSKGPRGETGAPRPGVEGPPGPPGAGEKSGPCADGAPAGAPGTPGQGTAGQGVVGLP 840
QY 625 GPPGERGPGSRGPPGADGVAGPKGPAERGSPGAPG-----KSGPGEAGRPGEAGLPGAK 681
Dd 841 GQRGEGFPGLPGPSGEPGKQPSGASGERGPPGPMGPPGLAGPPGSGREGAPGAGGSP 900
QY 682 GLTSPGSPGPDGKTGPPGAGQDGRPGPPGARGOAGVWPGPPGKGAAGEGKAGRG 741
Dd 901 GEDGSPGAKGDRGETGTPAGPPGAGKAGAPGVPAGKSGDRGETGTGAPGAPVGPAGAR 960
QY 742 GVPGPPGAVGAPGKDGE-----AGAQGPPG-----AGPAGERGEQ 777
Dd 961 GPAGPQGPGRDKGETGEGQDRGIKGHRGFSGLQGGPPGPPGSPGQSGSAGPAPRGP 1020
QY 778 GPAGSP---GFGGLPGAPGPGGAGKPGGQGVGDLGAPGSPGAG 820
Dd 1021 GSAGAPGKGLNGLPFGPIGPGPRGRTGDAGVGGPPGPPGPPG 1066

RESULT 8
US-09-548-608-18
; Sequence 18, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESS: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
```



```
QY 148 GAQPPGAGPAGGEGEQP-----ACSPGFQGLPGPAGPGEAGKGEQ 192
Db 192 GEPGPPGAGAGPAGNPGADGQCPAKGANGAPGAGPFGAGPFGSPGSPGPK 251
QY 193 GVPGLGAP-----GPSGA-----GPPGPTGLPGPPGERGGP 225
Db 252 GNSGEPGAPNKGDTGAKGEGCATGVQFPAGGEGKRGARGEPGSGLPFGPPGERGGP 311
QY 226 GSRFPAGADGAGKGPAGRGSPGAPGKSPGEGARPGAGLPAGAKGLTGSPPSGPPD 285
Db 312 GSRFPAGADGAGKGPAGRGSPGAPGKSPGEGARPGAGLPAGAKGLTGSPPSGPPD 371
QY 286 KTGPPGAGDGRPPGPPGARGQAGVMGFPKGAAGEPGKAGRGVPPGPAVGA 345
Db 372 KTGPPGAGDGRPPGPPGARGQAGVMGFPKGTAGEPGKAGEGLPFGPPGAVGA 431
QY 346 KKGEGAGQPPGAPGAGRGEGQAGSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAP 405
Db 432 KKGEGAGQAPGAPGAGRGEGQAGSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAP 491
QY 406 GPSGAPGPTGLPGPPGERGGSPGFPAGDGVAGKGPAGRGSPGAGKSPGEGA 465
Db 492 GPSGARGE-----RGFPGERGVQGP-----PGFAGPRGNNGAP 524
QY 466 GRPGEAGLPAGAKGLTGSPPGPPGQDGTGPPGAGQDGRPPGPPGARGQAGVMGFPK 525
Db 525 GN-----DGAKGDTGAPGAFGSQAGFLQGMPPGERGAAGLPKPKGRGDA-----GPK 572
QY 526 GAAGEPGKAGRGVPPGPAVGAUPGAKGEGAGCAQPPGAPGPA-----GERGGQAGSP 579
Db 573 GADGSPGKDGARGLTGPTGPPGAPGAPGKGEAGSPGPPGTGARGAPGDRGAGPPGA 632
QY 580 FQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAPGEPG-----TGUPGPPGSRGG---P 633
Db 633 GFAGPPGADGQPGAKGEGFDTGVKDGAGPPGAPGAPGPPGPNVNGAPGPKGPRGAAGPP 692
QY 634 GSRFPAGADGAGKGPAGRGSPGAPGKSPGEGARPGAGLPAGAKGLTGSPPSGPPD 693
Db 693 GATGPPGAAGRGVPPGSPGNAGPPGPPGVKGGKGRGETGAPGPPGVPVPPGPPGA 752
QY 694 KTGPPGAGDGRPPGPPGARGQAGVMGFPKGAAGEPGKAGRGVPPGPAVGA 753
Db 753 GEKSPGADGAGGFTGPPGAGVGLPQOR-----GERGFPGLP-----797
QY 754 KKGEGAGQPPGAPGAGRGEGQAGSPGFQGLPGPAG 793
Db 798 -----GPSGEPKQGPSGSSGERGPPGPMG 822

RESULT 10
US-08-468-996-10
; Sequence 10, Application US/08468996
; Patent No. 6645504
; GENERAL INFORMATION:
; APPLICANT: Weiner, Howard
; APPLICANT: Miller, Ariel
; APPLICANT: Zheng, Zheng
; APPLICANT: Ahmad, Al-Sabbagh
; TITLE OF INVENTION: BYSTANDER SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION
; FILE REFERENCE: 1010/16959-US3
; CURRENT APPLICATION NUMBER: US/08/468,996
; PRIOR FILING DATE: 2003-02-07
; PRIOR FILING DATE: 1992-02-28
; PRIOR FILING DATE: 1990-02-21
; PRIOR FILING DATE: 1990-02-21
; PRIOR FILING DATE: 1990-10-15
; PRIOR FILING DATE: 1987-06-24
; PRIOR FILING DATE: 1989-12-20
```

```
; PRIOR APPLICATION NUMBER: US 07/487,732
; PRIOR FILING DATE: 1990-03-02
; PRIOR APPLICATION NUMBER: US 07/551,632
; PRIOR FILING DATE: 1990-07-10
; PRIOR APPLICATION NUMBER: US 07/379,778
; PRIOR FILING DATE: 1989-07-14
; PRIOR APPLICATION NUMBER: US 07/507,826
; PRIOR FILING DATE: 1990-10-31
; PRIOR APPLICATION NUMBER: US 07/595,468
; PRIOR FILING DATE: 1990-10-10
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1017
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-08-468-996-10

Query Match 53.6%; Score 2488; DB 4; Length 1017;
Best Local Similarity 51.2%; Pred. No. 1.6e-153;
Matches 506; Conservative 49; Mismatches 265; Indels 168; Gaps 18;

QY 1 GPPGPPGTGLPGPPGERGGSPGSRGFPAGDGVAGKGPAGRGSPGAPGKSPGEGARGP 60
Db 25 GNPGEFGEVSGVSGMPGPRGPPGPKGDDGEAGKPKKAGRGPPGQAGRGFTGLP 84
QY 61 -----GEAGLPKAKGLTGP-----GSPGP-----DGKTGPPGPA--- 90
Db 85 GYKGRGVPLDGAKEAGAPGVKGESGSPGSGSPGMPGRLGPGERGTGPAGAGAR 144
QY 91 GQDGPAGPPGPPGARGQAGVMGFP-----GPKGAAGEPGKAGRGVPPGPAVGA 141
Db 145 GNDGQPPGAGPPGPPGPPGAGGPPGAPGAKGAGAGTARGGEGAGQGRGEGTGPSPGA 204
QY 142 -----GKDGEAGQPPGPPGAPGAGRGEGQAGSPGQGLPGPAGPPGEGAGKGEQGV 195
Db 205 GASGNPDTGTPGAKGSAGAPGAGPFPGRPPDPQATGTLGPKGQTKGFIAGFK 264
QY 196 GB---LGAPGSPGAPGEPGPTGLP-----GPPGERGGSPGSRGFPAGDGA 237
Db 265 GRQGKGEPPGAPGQAGPAGGEGKRGARGGPPGVGPIGPPGERGAPGNRGFPQDCLA 324
QY 238 GPKGAPGARGSPGAPGKSPGEGARPGFAGLPKAKGLTGSPPGPDGKTGPPGAPQD 297
Db 325 GPKGAPGARGSPGSLAGPKGANGDPGRGEPGLPARGLTGPPGAGPGRGKVPSPGAPGED 384
QY 298 GRPGGPPGPPGARGQAGVMGFPKGAAGEPGKAGRGVPPGPPGAVGAPGKDGEGAGAGPP 357
Db 385 GRPGGPPGQAGQPPGVGMPGPPGKAGGEPKAGEKGLPGAPGLRGLPGKDGEGAGPP 444
QY 358 GPAGPAGRGEGQAGSPGFPQGLPGPAGPGEAGKPGEGQVPGDLGAPGPPSPAGE --- 413
Db 445 GPAGPAGRGEGQAGSPGFPQGLPGPAGPGEAGKPGSDQGVGPPGAGAPGLVPPGERGFP 504
QY 414 -----PGPTGLP-----GPPGERGGSPGSRGFPAGDGVAGKGP 447
Db 505 GERGSPGAGQGLQGRKGLPTGTDTGPKCASGAPGPPGAGQPPGLQGMPPGERGAGIAGPK 564
QY 448 GERGSPGAPGKSPGEG-----AGRPGEAGLPKAKGLTGSPPSGSPGD 489
Db 565 GDRGVGKGPGEAGFPGKDGARGLTGPIGPPGPPAGANGKGEVGPFPGAGSAGARGPGE 624
QY 490 KTGPPGAGDGRPPGPPGARG-----GQAGVMGFPGPKGAAGBP-----GKA 534
Db 625 GETGPPGAPGAPGPPGADGQCPAKGEGQAGKDGAGAPGQPGSGAPGPPGCTGVTGPK 684
QY 535 GERGVPPGPPGA-----YGPAGKDGEGAGAGQPPGPA---GPAGRGEGQAGSPGFP 582
Db 685 GARGAQGPPGATGTPGAAAGRVGPPGSGNPNPVPGPSPGKDGKPKGARGDSGPPGRAGEP 744
QY 583 GLPGPAGPPGEGAGKPGEGQVPGDLGAPGSPGAPGPTGLPGPPGPPGSGSGSGFPAGD 642
Db 745 GLQGPAGPPGKGEPPGDDGPGSAGSPGPPQGLAQGRGLVGLPQGRGPPGLPSPGEP 804
```



QY 643 GVAGFGKGPAGSGSPGAGPKSGFGEAGRFGAELGPCA-----KGLT 684  
Db 805 GQCAPGASGDRGPPGVPGLTGPAGEFEREGSPGADGPPGRDGAAGVKGRGETGAV 864  
QY 685 GSPSGPDGKTGPPG-----AGQDRPFPFPFGARGAQVWGFPGPKGAAGEP 735  
Db 865 GAPCAPGPPGSPAGTGGKDRGEAGAOGFPGSPAGARGTGGQGGPRGDKGAGEP 924  
QY 736 GK---AGERGVPFGPGVAGPAGKDEAGAGQFPFGPAGPAGBERGQGPAGSPGFQGLPGPA 792  
Db 925 GERGLKHRRGFTGLQGLPGPPGSDQASGASGASGSPGPRFPFGVPGSGDKGANGTGPPI 984  
QY 793 GPPGEAKPGHQVPGDLGAPGSPGAG 820  
Db 985 GPPGRCRSGTGPAGPPGPNFGPPGPG 1012

## RESULT 11

US-08-642-255-62  
; Sequence 62, Application US/08642255  
; Patent No. 5773249  
; GENERAL INFORMATION:  
; APPLICANT: CAPPELLO, Joseph  
; APPLICANT: FERRARI, Franco A.  
; TITLE OF INVENTION: High Molecular Weight Collagen-Like  
; TITLE OF INVENTION: Protein Polymers  
; NUMBER OF SEQUENCES: 135  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/642,255  
; FILING DATE:

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: ROWLAND, Bertam I.  
REGISTRATION NUMBER: 20,015  
REFERENCE/DOCKET NUMBER: A55556-3/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299 FHT UR  
INFORMATION FOR SEQ ID NO: 62:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1084 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-642-255-62

Query Match 53.4%; Score 2478; DB 1; Length 1064;  
Best Local Similarity 53.8%; Pred. No. 7.5e-153;  
Matches 506; Conservative 33; Mismatches 281; Indels 120; Gaps 23;

QY 1 GPPGPGPTGLPQPERGGPSRGFPAGADGVAGPKGAPGERSGPP---AGPKSGPCEA 57  
Db 40 GPKGAGHAGPAGKAGPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFGSR 99  
QY 58 GRPGBAGLPGAKGLTSGSPGPDGKTGTPPGAGQDGRPPGPPGARGQAGVWGFPGPK 117  
Db 100 GPPGPPGAPGAPGSGRSDPFPAGPAGPAGPFGSRGDPFPFAGPAGPFGSRGDPFP 159

QY 118 GAAGEPKAGRCVPPGAVGPGAGKDEAGACQGP---PGPAGPAGERGE-----QGPA 168  
Db 160 GAPGFPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAGPAG 219  
QY 169 GSPGFGQGLPAGPAGPGEAGKPGEQGVPGDLGAPGSPG---AGEPGPTGLPFPFGG 225  
Db 220 GPPGSGRDPGPPGAGPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFG 279  
QY 226 GSRGPPGADGVAGPK---GPAGERSGPPAGPKGS---PGEAGRGEAGLPGAKGLTSP 279  
Db 280 GPAGPKGAGPAGPFGKAGHAGPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAG 339  
QY 280 GS-----PQDGKTGPPGAGQDGRGPPGPPGARGQAGVWGFPGPKGAAGEP 327  
Db 340 GAPGFPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAG 399  
QY 328 GXAGERGVPFGPAGVAGPAGKDEAGACQGP---PGPAGPAGERGE-----QGPA 378  
Db 400 GPPGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFG 459  
QY 379 GLPGPAGPGEAGKPGEQGVPGDLGAPGSPG---AGEPGPTGLPFPFGGSGSRFP 435  
Db 460 GPPGPPGAPGAPGPPGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFGSRGDP 519  
QY 436 GADGVAGPK-----QPAGERSGPPAGPKGSGFGEAGRP---GEAGLPGAKGLT 480  
Db 520 GAHGPAGPKGAGHAGPAGPKGAGHAGPAGPKGAGHAGPAGPKGAGHAGPAGPK 579  
QY 481 GSPGS---PQDGKTGPPGAGQDGRGPPGPPGARGQAGVWGFPGPKGAAGEP 537  
Db 580 GPPGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFG 639  
QY 538 GVPGPPGAVGAPAGKDEAGACQGP---PGPAGPAGERGE-----QGPA 588  
Db 640 GDFGPPGAPGAPGPPGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFGSRGDP 699  
QY 589 GPPGGEAGKPGEQGVPGDLGAPGSPG---AGEPGPTGLPFP-----PGR 630  
Db 700 GAPGFPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAG 759  
QY 631 GPPGSRGFP-----GADGVAGPK---GPAGERSGPPAGPKGSGFGEAGRP--- 672  
Db 760 GPPGSRGDPFPFAGHAGPAGPKGAGHAGPAGPKGAGHAGPAGPKGAGHAGPAG 819  
QY 673 GEAGLFGAKGLTSGPS---PQDGKTGPPGAGQDGRGPPGPPGARGQAGVWGFPGPK 729  
Db 820 GDFGPPGAPGAPGPPGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFGSRGDP 879  
QY 730 GAAGEPKAGRCVPPGAVGPGAGKDEAGACQGP---PGPAGPAGERGE-----QGPA 780  
Db 880 GAPGFPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAG 939  
QY 781 GSPGFGQGLPQAGPAGPGEAGKPGEQGVPGDLGAPGSPGAG 820  
Db 940 GPPGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFGSRGDPFPFAGPAGPFG 979

## RESULT 12

US-08-931-820-3  
; Sequence 3, Application US/08931820  
; Patent No. 6010863  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Assay for collagen degradation  
; NUMBER OF SEQUENCES: 4  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/931,820  
; FILING DATE:

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type II
US-08-931-820-3

Query Match          53.4%; Score 2476; DB 3; Length 1060;
Best Local Similarity 50.9%; Pred. No. 1e-152;
Matches 503; Conservative 45; Mismatches 272; Indels 168; Gaps 17;

QY 1 GPPGPGTGLPFPGERGSGRFGPGADGVAGPKGPAGERGSPGAPGKSPGEAGRP 60
DB 44 GNPGEPPGVSGPMRPPGPPGPKPGDDGEAGKPKAGRGPPGQARGFPTGLP 103
QY 61 -----GEAGLCAKLTGSP---GSPGP-----DGKTGPPGPA--- 90
DB 104 GVKGHRGVPGLDGAKEAGAGVKGESGPGNGSPGMRGLPGBRGRTGAGAAAR 163
QY 91 QDGRGPPGPPGARGQAGVMGFP-----GPKGAAGBPCKAGRGVPPGPAVGPA 141
DB 164 GNDGQPGAGPPGVPVAGGSGFFGAPCAKGEAGTCAKGPEGAQGRGPPGSPGPA 223
QY 142 -----GKDGAGAGQPPGPPAGPAGERGEGQAPGSPFQGLPGPAGPPGAGRGPGGV 195
DB 224 GASGNPFGTDGTPGAKSGAGFAGAPFGPGRGPPDQATGFLGPKGQTGKPGIAGFK 283
QY 196 GD---LGAPGSPGAGBPPTGLP-----GPPBERGGSGRGGPPGADGVA 237
DB 284 GEQPKGEFPAGPQAGPAGPAGESGKRGARCEPGCVGPIGPPGERGAFNGRPPQDGLA 343
QY 238 GPKGPAERGPSPGAPGKSPGEGAGRGEGAGLPAKAGLTGSPGSPGPDGNTGPPGAGQD 297
DB 344 GPKGAPGERSGLAGPKGANGDPERPCEPGLPGARGLTGRPDAGPQGVKVPSPGAPGD 403
QY 298 GRPGPPGPPGARGAGVMPGPPGAAAGEPKKAGRGVPPGPAVGPAGKDGAGAGGPP 357
DB 404 GRPFPFGQARGQGVWVFPFGPKANGEPKAGKGLPGAPLURLPGKDGETGAEPPP 463
QY 358 GPAGPAGERGQGPAGSPFGGLPGPAGPGEAGKPGEGQGVGPDGLGAPGSPGAGE--- 413
DB 464 GPAGPAGERGQGPAGSPFGGLPGPAGPGEAGKPGEGQGVGPDGLGAPGSPGAGE--- 523
QY 414 -----PGTGLP-----GPPGERGSGRPPPGADGVAGPKGPA 447
DB 524 GERGSPGAQLQGPRGLPFGTPTDGPCKAGSPAGPPQAQPPGLQGMPPGERGAAGTAPK 583
QY 448 GERGS-----PAPAGKSPGEGAGRPGEAGLPAGKAGLTGSP 483
DB 584 GDRGDVGEKGPAGPKDGGRLTGP:GPPCPAGANGKEGVEGPPGPPAGSAGARGAPGER 643
QY 484 GSPGPDGKTGPPGAGQDGRGPPGPPGARGQAGVWGMFGPKKAAAEPP-----GKA 534
DB 644 GETGPPGTSIAGPFGADGQFCAKGEQGEAGQKGDAGAPGQGPSPGAPGQGTGTGTPK 703
QY 535 GERGVPPGPA-----VGPAGKDGAGAGAGPPGPA---GPAGERGQGPAGSPGFQ 582
DB 704 GARGAQFPFGATGFGAAGRVGPPGSGNPNPFPFPFPCKDGPCKAGDSGPPGAGERP 763
QY 583 GLPFPAGPPGAGKPGGQGVGDLGAPGSPGAPGEPGPTGLPFPGERGSGRGPFGAD 642
DB 764 GLQFAGPPGPKGFPDGGSGABPPGQGLAGQRGIVLPGQRGRRGPPGLPGPSGEP 823

```

```

QY 643 GVAGPKPAGERGSPGAPGKSPGEGAGRPGEAGLPGA-----KGLT 684
DB 824 GQAGPAGSGDRGPPGPPGVPGLTGPAGEPPGREGSPGADGPPRDRGAAAGVKDRGETCAV 883
QY 685 GSPGSPGDPGKTGPPGP-----AQDGRPPGPPGARGQAGVWGFPGPKAAAEPP 735
DB 884 GAPGAPGPPGSPGAPGTPKQKDRGEAGAAQGMGSPGAGARGIQGPGQPRGDKGEAGEP 943
QY 736 GK---AGERGVPPGPAVGPAGKDGEGAGAAQAPPPGAPGAGERGEGQSPGSPFQGLPGPA 792
DB 944 GERGLKGRHGFGLQGLFPGPPGSDQAGSPGSPGPPGPPGVPGPSKDGANGIPGPI 1003
QY 793 GPPGAGRGKPGEGQVPGDILGAPGSPGAP 820
DB 1004 GPPGPRGRSGETGPAGPFGNPGPPGPPG 1031

RESULT 13
US-08-963-825-20
; Sequence 20, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963.825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187.319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogozis, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
US-08-963-825-20

Query Match          53.4%; Score 2476; DB 3; Length 1418;
Best Local Similarity 50.9%; Pred. No. 1.3e-152;
Matches 503; Conservative 45; Mismatches 272; Indels 168; Gaps 17;

QY 1 GPPGPGTGLPFPGERGSGRFGPGADGVAGPKGPAGERGSPGAPGKSPGEAGRP 60

```

Db 156 GNPGEPPGVSGVNGPRGPPGPKGDDGEGAKPKGKAGERGPPGQAGRGFTGFLP 215  
Qy 61 -----GBAGLPGAKGLTGSP-----GSPGP-----DGKTPPGPA--- 90  
Db 216 GVKHGRGVGLDGAKEAGAPGVKSGSGSPGPMGRGLPGRGRTGAGAGAR 275  
Qy 91 QODRRPPTPPGARGQAGVGMFP-----GPKGAAGEPKAGRGVPGPPGAVGA 141  
Db 276 GNDQGPAGPPGVGAGGPPGAGPAGKAGAGTARGPAGAGGPPGPPGPPGPA 335  
Qy 142 -----GKDGAGAGQPPGAGPAGRGEGQAGPAGPAGPAGPAGPAGPAGPAG 195  
Db 336 GASGNFTGIPGAKGSAGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 395  
Qy 196 GD---LCAPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 237  
Db 396 GEQGPKEGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 455  
Qy 238 GPKPAGRGSPGAGPKSGPGEAGRPCEAGLPKAGLGTGSPSPGPPGPPGPPGAG 297  
Db 456 GPKAPGERSGLAGPKGANGDPRGEPGLPGARGLTGPPGPDAGPQGVGSGAG 515  
Qy 298 GRPPTPPGARGQAGVNGPPGPKGAAGEPKAGRGVPPGPPGAVGAGKDGAGAG 357  
Db 516 GRPPTPPGARGQAGVNGPPGPKGAAGEPKAGRGVPPGPPGAVGAGKDGAGAG 575  
Qy 358 GPAPAGRGSPGAGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 413  
Db 576 GPAPAGRGSPGAGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 635  
Qy 414 -----PGTGLP-----GPPGERGPPGPPGPPGPPGPPGPPGPPGPPG 447  
Db 636 GERGSPGAGLQGRGLPGTGTDPGPKGASGPPGPPGPPGPPGPPGPPGPPG 695  
Qy 448 GERGS-----PPAPGKSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 483  
Db 696 GDRGVGKPEGAPGKDGGRGLTGPIGPPGPPGPPGPPGPPGPPGPPGPPG 755  
Qy 484 GSPGPDGTPGPPGAGQDGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 534  
Db 756 GETGPPGTSIAGPPGADGQFGAKGEQGEAQKGDAGAPGPPGPPGPPGPPG 815  
Qy 535 GERGVGPPGPA-----VGPAGKGEAGAGPPGPPA-----GPAGRGPPG 582  
Db 816 GARGAGPPGATGTFPGAAGRVGPPGPPGPPGPPGPPGPPGPPGPPGPPG 875  
Qy 583 GLPGPAGPPGAGKRGEGVGDGLGAPGSPGPPGPPGPPGPPGPPGPPGPPG 642  
Db 876 GLQGPAGPPGKGEPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 935  
Qy 643 GVAGPKPAGRGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 684  
Db 936 GQOGAPGASGDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 995  
Qy 685 GSPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 735  
Db 996 GAPGAPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1055  
Qy 736 GK---AGRGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 792  
Db 1056 GERGLKHRGFTGLQGLPQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1115  
Qy 793 GPPGAGKRGEGVGDGLGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 820  
Db 1116 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1143

RESULT 14  
US-09-010-999-1  
; Sequence 1, Application US/09010999  
; Patent No. 6132976  
; GENERAL INFORMATION:

APPLICANT: Poole, Anthony R.  
APPLICANT: Hollander, Anthony P.  
APPLICANT: Billingshurst, R.C.  
TITLE OF INVENTION: IMMUNOSAYS FOR THE MEASUREMENT OF  
TITLE OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/010,999  
FILING DATE: 22-JAN-1998  
CLASSIFICATION: 4335  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/448,501  
FILING DATE: 17-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/984,123  
FILING DATE: 04-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Best, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 032931/0212  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1418 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE: Human Type II Collagen  
ORGANISM: Human Type II Collagen  
US-09-010-999-1

Query Match 53.4%; Score 2476; DB 3; Length 1418;  
Best Local Similarity 50.9%; Pred. No. 1.3e-152;  
Matches 503; Conservative 45; Mismatches 272; Indels 168; Gaps 17;  
Qy 1 GPPGEPGPTGLPQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 60  
Db 156 GNPGEPPGVSGVNGPRGPPGPKGDDGEGAKPKGKAGERGPPGQAGRGFTGFLP 215  
Qy 61 -----GKAGLPGAKGLTGSP-----GSPGP-----DGKTPPGPA--- 90  
Db 216 GVKHGRGVGLDGAKEAGAPGVKSGSGSPGPMGRGLPGRGRTGAGAGAR 275  
Qy 91 QODRRPPTPPGARGQAGVGMFP-----GPKGAAGEPKAGRGVPPGPPGAVGA 141  
Db 276 GNDQGPAGPPGVGAGGPPGAGPAGKAGAGTARGPAGAGGPPGPPGPPGPA 335  
Qy 142 -----GKDGAGAGQPPGAGPAGRGEGQAGPAGPAGPAGPAGPAGPAGPAG 195  
Db 336 GASGNFTGIPGAKGSAGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 395  
Qy 196 GD---LCAPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 237  
Db 396 GEQGPKEGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 455  
Qy 238 GPKPAGRGSPGAGPKSGPGEAGRPCEAGLPKAGLGTGSPSPGPPGPPGPPGAG 297



Search completed: October 18, 2004, 13:44:02  
Job time : 39.5431 secs

*This Page Blank (usp10)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:42:06 ; Search time 105.622 Seconds  
(without alignments)  
2513.149 Million cell updates/sec

Title: US-10-658-989A-3  
Perfect score: 4640  
Sequence: 1 GPGEGPTGLPGPPGREGG.....GEOGVFDLGAQPSGAGG 821

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1360919 seqs, 323318874 residues

Total number of hits satisfying chosen parameters: 1360919

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2765.5	59.6	1449	15	US-10-402-089-8
2	2765.5	59.6	1449	15	US-10-402-072A-8
3	2765	59.6	1057	15	US-10-104-889-20
4	2765	59.6	1107	15	US-10-104-889-11
5	2765	59.6	1171	15	US-10-104-889-8
6	2765	59.6	1388	15	US-10-104-889-10
7	2765	59.6	1464	14	US-10-216-705-21
8	2765	59.6	1464	16	US-10-357-851-1
9	2765	59.6	1464	16	US-10-358-024-1
10	2765	59.6	1464	17	US-10-788-792-150
11	2761	59.5	1461	16	US-10-468-091-25
12	2761	59.5	1464	14	US-10-291-265-243
13	2759	59.5	1057	15	US-10-104-889-16
14	2755	59.4	1464	10	US-09-918-715-261

Sequence 159, Appl  
Sequence 36, Appl  
Sequence 2, Appl  
Sequence 65, Appl  
Sequence 28, Appl  
Sequence 78, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 6, Appl  
Sequence 26, Appl  
Sequence 18, Appl  
Sequence 49, Appl  
Sequence 10, Appl  
Sequence 1, Appl  
Sequence 20, Appl  
Sequence 5, Appl  
Sequence 48, Appl  
Sequence 6, Appl  
Sequence 12, Appl  
Sequence 12, Appl  
Sequence 4, Appl  
Sequence 6, Appl  
Sequence 6, Appl  
Sequence 228, Appl  
Sequence 68, Appl  
Sequence 33, Appl  
Sequence 72, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 103, Appl

US-10-060-036-159  
US-10-171-311-36  
US-10-149-352-2  
US-10-177-293-65  
US-10-301-822-28  
US-10-734-564-79  
US-10-402-089-2  
US-10-402-072A-2  
US-10-104-889-6  
US-10-468-091-26  
US-10-058-124-18  
US-10-342-331-49  
US-10-639-286-10  
US-10-194-441A-1  
US-10-058-124-20  
US-10-468-091-5  
US-10-134-441A-48  
US-10-468-091-6  
US-10-402-089-12  
US-10-402-072A-12  
US-10-402-089-4  
US-10-402-089-6  
US-10-402-072A-4  
US-10-402-072A-6  
US-09-918-715-226  
US-10-177-293-68  
US-10-301-822-33  
US-10-257-021-72  
US-10-357-851-3  
US-10-358-024-3  
US-10-734-564-103

ALIGNMENTS

RESULT 1  
US-10-402-089-8  
; Sequence 8, Application US/10402089  
; Publication No. US20040005663A1  
; GENERAL INFORMATION:  
; APPLICANT: Bell, Marcum P.  
; APPLICANT: Neff, Thomas B.  
; APPLICANT: Polarek, James W.  
; APPLICANT: Seeley, Todd W.  
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS  
; FILE REFERENCE: PP0402.3 CON  
; CURRENT APPLICATION NUMBER: US/10402,089  
; CURRENT FILING DATE: 2003-03-26  
; PRIOR APPLICATION NUMBER: US 09/709,700  
; PRIOR FILING DATE: 2000-11-10  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 8  
; LENGTH: 1449  
; TYPE: PRT  
; ORGANISM: Sus scrofa  
; US-10-402-089-8

Query Match 59.6%; Score 2765.5; DB 15; Length 1449;  
Best Local Similarity 56.7%; Pred. No. 4.5e-131;  
Matches 559; Conservative 32; Mismatches 226; Indels 169; Gaps 21;  
  
QY 1 GPGEGPTGLPGPPGREGG-----RGFPGA-----DGVAGPKGAGERSGPGA 48  
Db 193 GPGAGPGQFGPGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 252  
QY 49 GPKGSPEAGRP-----GEAGLFGAGLGTSGSPGPPGDKTGPFGPAGD 93  
Db 253 GARGLPGTAGLPGMKHRCFSGLDGAKGDAGPAGPKGPGSGENGAPQMGPRGLPGR 312  
QY 94 GRPGPPGPGAGGQ-----AGMGFPKGAAGB---PGKAGRGVPGPPGAVGPKDGA 147

Db 313 GRPFPAGARGNDGATGAAGPPGPTGAPGPPFPGAVGAKGAGPQAGSGEPQGV 372  
Qy 148 GAQGGPPGAGPAGERGEGP-----AGSPGQGLPGGAPPGPGEAGKEQ 192  
Db 373 GEPGPPGAGAGPAGNPGADQCGKAGAPGAGPAGPAGPAGPAGPAGPAGP 432  
Qy 193 GVPGLGAP-----GFSGPA-----GEPGPTGLPGPPGERGGP 225  
Db 433 GNSGEPGAPGKGTGAKGEPGPTGVQGPAGPAGGEGKRGARGGPPAGLPGERGGP 492  
Qy 226 GSRFPAGADGAVGKPRAGERGSGPAGPKSPGAEAGPGEAGLPKAKGLTSGPSGPPD 285  
Db 493 GSRFPAGADGAVGKPRAGERGSGPAGPKSPGAEAGPGEAGLPKAKGLTSGPSGPPD 552  
Qy 286 KGTGPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGBPAGKAGRGVPPGPPGVA 345  
Db 553 KGTGPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGBPAGKAGRGVPPGPPGVA 612  
Qy 346 KGBAGAGQPPGAGPAGERGEGPAGSPGQGLPGAPPPGAGPAGPAGPAGPAG 405  
Db 613 KGBAGAGQPPGAGPAGERGEGPAGSPGQGLPGAPPPGAGPAGPAGPAGPAG 672  
Qy 406 GPSGPA-----PGP-----TGLPGERGPPGSGRGP 435  
Db 673 GPSGARGPFPGERGVQGPAGPRGANGAPGNDGAKGADGAPGAGPAGPAGPAG 732  
Qy 436 GADGAVGKPRAGERGSGPAGPKSPGAEAGPGEAGLPKAKGLTSGPSGPPD 486  
Db 733 GERGAAGLPKGRDGDAGPKADGAPKGDVRLGTGPIGPPGAGPAGPAGPAGPAG 792  
Qy 487 GPDGKTGPPGAGQDGRPP-----PGPCARGQAGVMGF-----PGPKAAGBPAGKAGRGV 539  
Db 793 GPTGARGAPGDRGEPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 852  
Qy 540 PGPGA-----VGPAGKDGAGAGQPPGPPAGPAGE-----RGEQPPAGSPGQGLPGP 1089  
Db 759 -----ACAGQPPGP-----AGPAGERGEGPAGSP-----GFGGLPGPAGP 794  
Db 1090 RGIKGRGFSGLQPPPPGSGPQSGASGAPGAPGPPGSGAGAPKDGGLNGLPPIGP 1149  
Qy 795 PGEAGKPGEGVPGDLGAPGSGPAG 820  
Db 1150 PGPGRGTGAGVPGPPGPPGPPGPPG 1175

## RESULT 2

US-10-402-072A-8

; Sequence 8, Application US/10402072A

; Publication No. US20040018592A1

; GENERAL INFORMATION:

; APPLICANT: Bell, Marcum P.

; APPLICANT: Neff, Thomas B.

; APPLICANT: Polarek, James W.

; APPLICANT: Slesley, Todd W.

; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS

; FILE REFERENCE: FPO402.2 CON

; CURRENT FILING DATE: 2003-03-26

; PRIOR APPLICATION NUMBER: US 09/709,700

; PRIOR FILING DATE: 2000-11-10  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 8  
; LENGTH: 1449  
; TYPE: PRT  
; ORGANISM: Sus scrofa  
US-10-402-072A-8

Query Match 59.6%; Score 2765.5; DB 15; Length 1449;

Best Local Similarity 56.7%; Pred. No. 4.5e-131; Indels 169; Gaps 21;

Matches 559; Conservative 32; Mismatches 226

Qy 1 GPPGPGPTGLPGPPGERGGPQS-----RGFPGA-----DGVAGPKPAGERGSGP 48  
Db 193 GPPGAPGPGFQGGPGEPEGASGPMGPRGPPGPPGKNGDDGAGKPRGPRGPPG 252  
Qy 49 GPKGSPGAGRP-----CEAGLPKAKGLTSGPSGPPGDKTGPAGPAG 93  
Db 253 GARGLPCTAGLPCKMGRHGFSLDGAKDAGPAGPKGPPSGENGAPGQMPGLPGR 312  
Qy 94 GRPFPAGPAGRGNDGATGAAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 147  
Db 313 GRPFPGAGARGNDGATGAAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 372  
Qy 148 GAQGGPPGAGPAGERGEGP-----AGSPGQGLPGPAGPAGPAGPAGPAG 192  
Db 373 GEPGPPGAGAGPAGNPGADQCGKAGAPGAPGPPGARGSPGSGPPGPK 432  
Qy 193 GVPGLGAP-----GPSGPA-----GEPGPTGLPGPPGERGGP 225  
Db 433 GNSGEPGAPGKGTGAKGEPGPTGVQGPAGPAGGEGKRGARGGPPAGLPGERGGP 492  
Qy 226 GSRFPAGADGAVGKPRAGERGSGPAGPKSPGAEAGPGEAGLPKAKGLTSGPSGPPD 285  
Db 493 GSRFPAGADGAVGKPRAGERGSGPAGPKSPGAEAGPGEAGLPKAKGLTSGPSGPPD 552  
Qy 286 KGTGPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGBPAGKAGRGVPPGPPGVA 345  
Db 553 KGTGPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGBPAGKAGRGVPPGPPGVA 612  
Qy 346 KGBAGAGQPPGAGPAGERGEGPAGSPGQGLPGAPPPGAGPAGPAGPAGPAG 405  
Db 613 KGBAGAGQPPGAGPAGERGEGPAGSPGQGLPGAPPPGAGPAGPAGPAGPAG 672  
Qy 406 GPSGPA-----PGP-----TGLPGERGPPGSGRGP 435  
Db 673 GPSGARGPFPGERGVQGPAGPRGANGAPGNDGAKGADGAPGAGPAGPAGPAG 732  
Qy 436 GADGAVGKPRAGERGSGPAGPKSPGAEAGPGEAGLPKAKGLTSGPSGPPD 486  
Db 733 GERGAAGLPKGRDGDAGPKADGAPKGDVRLGTGPIGPPGAGPAGPAGPAGPAG 792  
Qy 487 GPDGKTGPPGAGQDGRPP-----PGPCARGQAGVMGF-----PGPKAAGBPAGKAGRGV 539  
Db 793 GPTGARGAPGDRGEPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 852  
Qy 540 PGPGA-----VGPAGKDGAGAGQPPGPPAGPAGE-----RGEQPPAGSPGQGLPGP 587  
Db 853 AGPPGATGFPGAAGRVGPPGSGNAGPPGPPGARGKSGKPRGETGPPAGRPGEAGP 912  
Qy 588 AGPGEAGKPGEGVPGDLGAPGSGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 647  
Db 913 PGPAGEKSPADGAPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 972  
Qy 648 KGPAGERGSPAGPKSPGAEAGPGEAGLPKAKGLTSGP-----GSPGPD-----GKTGPPGP 701  
Db 973 SGPSGERGPPGPMGF-----PGLAGPFBGREGAPGAGSGPCRDGAPGPKDRESG 1029  
Qy 702 AGQDGRPFPAGPAGQAGVMGPPGKGAAGBPAGKAGRGVPPGPPGAVGAPKAG 758  
Db 1030 PGAPGAPGVPVAGPKSGDRGTGTGAPGAPGVPVAGPAGPAGPAGPAGPAGPAG 1089  
Qy 759 -----ACAGQPPGP-----AGPAGERGEGPAGSP-----GFGGLPGPAGP 794  
Db 1090 RGIKGRGFSGLQPPPPGSGPQSGASGAPGAPGPPGSGAGAPKDGGLNGLPPIGP 1149  
Qy 795 PGEAGKPGEGVPGDLGAPGSGPAG 820  
Db 1150 PGPGRGTGAGVPGPPGPPGPPGPPG 1175







Query Match 59.6%; Score 2765; DB 15; Length 1171;  
Best Local Similarity 55.4%; Pred. No. 4e-131;  
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPGEPCPTGLPDPGERGSGRPFPGADGVAGKGPAGRGSPGAPGKSGPGEA--- 57  
DB 18 GPMGSGPRGLPFGPCAPQFQGGPFGEFGEFGASGPMGPRGPGKNGDDGEAGKP 77  
QY 58 GRPGEAGLP---GAKGLTSGSPSGP-----DGKTGPPGAGQDGRPGPPPGCAR 105  
DB 78 GRPGERGPPGOGARGLPCTAGLPGMKHGRGSGLDGAKDAGPAGKGEFGSPGNGAP 137  
QY 106 GQAGVNGFFPGKGAAGEPKKAGER-----GVPGPPGAVGPA-----GKDGAGAG 150  
DB 138 GQMGPRGLPGERGPRGAPGAGNDGATGAAGPPTGAPGPPGPAVAGKAGAGPQ 197  
QY 151 GPPGAPGAPAGERGEGP-----AGSPGFQGLPGPA 180  
DB 198 GRPGEPPGVGRGPPPPGAPGAGNPGADGQPGAKGANGAPGAPGPPGARGPS 257  
QY 181 GPPGEAGKGEQVPGDLGAP-----GPSGPA-----GSPGPT 213  
DB 258 GPOGPGGPPGPKNSGEPGAPGSKDGTAKGEPGVPVQPPGAPGEGKRGARGPPT 317  
QY 214 GLPDPGERGSGRPFPGADGVAGKGPAGRGSPGAPGKSGPGEAGLPAGK 273  
DB 318 GLPDPGERGSGRPFPGADGVAGKGPAGRGSPGAPGKSGPGEAGLPAGK 377  
QY 274 GLTSPGSPGPDGKTGPPGAPQDGRPPGPPGARGQAGVNGFFGPKGAAGEPKKAGER 333  
DB 378 GLTSPGSPGPDGKTGPPGAPQDGRPPGPPGARGQAGVNGFFGPKGAAGEPKKAGER 437  
QY 334 GVPFPFAGVAGPKGDEAGAGQAPGAPGAPRGSGPAGKSGPGEAGRGAGLPAGK 393  
DB 438 GVPFPFAGVAGPKGDEAGAGQAPGAPGAPRGSGPAGKSGPGEAGRGAGLPAGK 497  
QY 394 GEQGVPGDLGAPGSPGAGE-----PGP-----TCLPDP 423  
DB 498 GEQGVPGDLGAPGSPGAGERGFPGERGVQPPGAPRGANGAPGNDGAKDAGAPAG 557  
QY 424 GERGGSPGKFPFAGVAGPKGAPGAPRGSPGAPGKSGPGE-----AGRPEAGLP 474  
DB 558 GSQAGPLQGMGERGAAGLPKDGDRDAGPKGADGSPGKDGVRGLTGPITGPPGAPAG 617  
QY 475 GAK-----GLTSPGSPGPDGKTGPPGAPQDGRPPGPPGARGQ----- 515  
DB 618 GDKSGSPGAPGPTGARGAPGDRGEPGPPGAPGAPGPPGAGPQFQKAGKEDGAKGDA 677  
QY 516 -----AGVMGFPKGAAGEPKKAGERGVPPGPA-----VGPAGKDGAGAGAG 561  
DB 678 GPPGAPGAPGPPGPIGNVAPCAKAGSAGPPGATGPPGAGRVGPPGSGNAGPPGPP 737  
QY 562 GPAGP---AGERGEGPAGSPGQGLPAPGPPGAGKPGGQGVPGDLGAPGSPGAPGEP 618  
DB 738 GPAGKEGKGRGETGAPRPGVEVPPGPPGAPGKSGPAGDGFAGAPGTGPPQIAGQR 797  
QY 619 GTGLPFPGERGGPSRPFPCADGVAGKGPAGRGSPGAPG---KGSPEAGRPGEA 675  
DB 798 GVVGLPQGEGERGFPGLPSPGSEPKQKPSGASGERGPPGPPGLAGPPGSEGREAP 857  
QY 676 GLPAGKGLTSGSPGPPDGKTGPPGAPQDGRPPGPPGARGQAGVNGFFGPKGAAGEP 735  
DB 858 GAEGSPFRDGSFCAKDRGETGAPGPPGAPGAPGPPVPGPAGKSGDRGETGAPGAPV 917  
QY 736 GKAGERGVPPGAVGAPKAGE-----AGAGPPG-----AGPA 771  
DB 918 GPAGARGPAGPPGPDGKETGEQDGRGKIHGRGSGLGPPGPPGSPGEGQPSGASGPA 977  
QY 772 GERGQGPAGSP---GFQGLPAGPAGGAGKPGGQGVPGDLGAPGPPSGPAG 820  
DB 978 GPRGPPGSAGAPGKDLNGLPGLPIGPPGRGRTGAGVPGPPGPPGPPG 1029

RESULT 6  
US-10-104-889-10  
; Sequence 10, Application US/10104889  
; Publication No. US20040086961A1  
; GENERAL INFORMATION:  
; APPLICANT: GRUSKIN, ELLIOT A.  
; BUECHTER, DOUGLAS  
; BROKAW, JANE  
; ZHANG, GUANGHUI  
; PAOLELLA, DAVID  
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DILWORTH & BARRESE  
; STREET: 333 EARLE OVINGTON BOULEVARD  
; CITY: UNIONDALE  
; STATE: NY  
; COUNTRY: U.S.A.  
; ZIP: 11553  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10104,889  
; FILING DATE: 22-Mar-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/169,768  
; FILING DATE: 09-OCT-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEEN, JEFFREY S  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (516) 228-8484  
; TELEFAX: (516) 228-8516  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1388 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-10-104-889-10

Query Match 59.6%; Score 2765; DB 15; Length 1388;  
Best Local Similarity 55.4%; Pred. No. 4.6e-131;  
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPGEPCPTGLPDPGERGSGRPFPGADGVAGKGPAGRGSPGAPGKSGPGEA--- 57  
DB 18 GPMGSGPRGLPFGPCAPQFQGGPFGEFGEFGASGPMGPRGPPGKNGDDGEAGKP 77  
QY 58 GRPGEAGLP---GAKGLTSGSPSGP-----DGKTGPPGAPQDGRPGPPPGCAR 105  
DB 78 GRPGERGPPGOGARGLPCTAGLPGMKHGRGSGLDGAKDAGPAGKGEFGSPGNGAP 137  
QY 106 GQAGVNGFFPGKGAAGEPKKAGER-----GVPGPPGAVGPA-----GKDGAGAG 150  
DB 138 GQMGPRGLPGERGPRGAPGAGNDGATGAAGPPTGAPGPPGPAVAGKAGAGPQ 197  
QY 151 GPPGAPGAPAGERGEGP-----AGSPGFQGLPGPA 180  
DB 198 GRPGEPPGVGRGPPPPGAPGAGNPGADGQPGAKGANGAPGAPGPPGARGPS 257  
QY 181 GPPGEAGKGEQVPGDLGAP-----GPSGPA-----GSPGPT 213  
DB 258 GPOGPGGPPGPKNSGEPGAPGSKDGTAKGEPGVPVQPPGAPGEGKRGARGPPT 317  
QY 214 GLPDPGERGSGRPFPGADGVAGKGPAGRGSPGAPGKSGPGEAGLPAGK 273

Db 318 GLPSPGRRGGSGFTGADGVAGPKGAPGAGERSGGPAGKSGSGRAGRPGEAGLPGAK 377  
 QY 274 GLTSGSPGPDGKTPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEFGKAGER 333  
 Db 378 GLTSGSPGPDGKTPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEFGKAGER 437  
 QY 334 GVPSPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 393  
 Db 438 GVPSPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 497  
 QY 394 GEGVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 423  
 Db 498 GEGVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 557  
 QY 424 GERGGSGRFFGADGVAGPKGAPGAGERSGGPAGKSGSGRAGRPGEAGLPGAK 474  
 Db 558 GSGGAPGLQMPGGERGAAGLPGPKDGRDAGPKGADGSGDKGVRLTGTPTGPPGAPAG 617  
 QY 475 GAK-----GLTSGSPGPDGKTPPGAGQDGRPPGPPGARGQ-----515  
 Db 618 GDXGESGSPGAPGTGARGAPDGRGEPGPPGAPGAGPPGADGQPGKAGEFGDAGAKGDA 677  
 QY 516 -----AGVMGPPGKGAAGEFGKAGERGVPQPPGA-----VGPAGKDGAGAGAGAG 561  
 Db 678 GPPGAPGAGPPGPIGNVAGPKAGARGSGAGPPGATGFFGAAGRVGPPGSGNAGPPGPP 737  
 QY 562 GPAGP---AGERGEGSPGSGFQGLPGAPGPPGAGKSGSGRAGRPGEAGLPGAK 618  
 Db 738 GPAGKGGKGRGETGAPRGEVGPVPPGAPGKSGFGADGAPAGETGPPGAGIAGQR 797  
 QY 619 GPTGLPSPGSGRGGSGRFFGADGVAGPKGAPGAGERSGGPAGP---KXSGEAGRPGEA 675  
 Db 798 GVVGLPQGRGERFFGLPSPGSEPKQSGASGERGPPGPPGGLAGPPGSGSGREGAP 857  
 QY 676 GLPAGKGLTSGSPGPDGKTPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEFG 735  
 Db 858 GAEGSPGSGRGGSGRGGTGTGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 917  
 QY 736 GXAGERGVPPGAVGAPGAGKDG-----AGAGGPPG-----AGPA 771  
 Db 918 GPAGARGPAGPPGPDGKGTGEGQDRIKGRHSGSLGPPGPPGSGEGSPGASGSPA 977  
 QY 772 GERGGQGPAGSP---GFOGLPGAPGAGKSGRGGVGGDLGAPGSGGAPG 820  
 Db 978 GPRGPPGSGAGPKDGLNGLPGIPGPPGRGTGADGVVGPVPPGPPGPPG 1029

## RESULT 7

US-10-216-705-21  
 ; Sequence 21, Application US/10216705  
 ; Publication No. US20030096973A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Meristem Therapeutics, S.A.  
 ; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me  
 ; TITLE OF INVENTION: Obtaining Such and Their Uses  
 ; FILE REFERENCE: 1149-3 DIV  
 ; CURRENT APPLICATION NUMBER: US/10/216,705  
 ; CURRENT FILING DATE: 2002-08-09  
 ; PRIOR APPLICATION NUMBER: US 09/331,347  
 ; PRIOR FILING DATE: 1999-08-17  
 ; NUMBER OF SEQ ID NOS: 22  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 21  
 ; LENGTH: 1464  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-216-705-21

Query Match 59.6%; Score 2765; DB 14; Length 1464;  
 Best Local Similarity 55.4%; Pred. No. 4,8e-131;  
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGEGPTGLPSPGRRGGSGRFFGADGVAGPKGAPGAGERSGGPAGKSGSGRAGRPGEAGLPGAK 57

Db 179 GVMGSGPRKGLPGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 238  
 QY 58 GRPGEAGLP---GAKGLTSGSPGPP-----DGKTPGPPGAGQDGRPPGPPGARG 105  
 Db 239 GRPGERGPPGQAGRLPGTAGLPGMKHGRGSLDGAAGAGAGAGAGAGAGAGAGAGAG 298  
 QY 106 QGAGVMGPPGPPGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 150  
 Db 299 GQNGPRLGPGERGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 358  
 QY 151 GPPGAPGAPGAGERGGP-----AGSPGFGQLPGPA 180  
 Db 359 GPRGSGPPGVGPPGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 418  
 QY 181 GPPGEAGKFGGQVPGDLGAP-----GPSGA-----GEPGT 213  
 Db 419 GPQPGPPGPPGKNGSGEPGAPGSKDGTAKGPPGVVQVPPGPPGAGBEGKRGARGEPT 478  
 QY 214 GLPSPGSGRGGSGRFFGADGVAGPKGAPGAGERSGGPAGKSGSGRAGRPGEAGLPGAK 273  
 Db 479 GLPSPGSGRGGSGRFFGADGVAGPKGAPGAGERSGGPAGKSGSGRAGRPGEAGLPGAK 538  
 QY 274 GLTSGSPGPDGKTPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEFGKAGER 333  
 Db 539 GLTSGSPGPDGKTPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEFGKAGER 598  
 QY 334 GVPSPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 393  
 Db 599 GVPSPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 658  
 QY 394 GEGVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 423  
 Db 659 GEGVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 718  
 QY 424 GERGGSGRFFGADGVAGPKGAPGAGERSGGPAGKSGSGRAGRPGEAGLPGAK 474  
 Db 719 GSGAPGLQMPGGERGAAGLPGPKDGRDAGPKGADGSGDKGVRLTGTPTGPPGAPAG 778  
 QY 475 GAK-----GLTSGSPGPDGKTPPGAGQDGRPPGPPGARGQ-----515  
 Db 779 GDXGESGSPGAPGTGARGAPDGRGEPGPPGAPGAGPPGADGQPGKAGEFGDAGAKGDA 838  
 QY 516 -----AGVMGPPGKGAAGEFGKAGERGVPQPPGA-----VGPAGKDGAGAGAGAG 561  
 Db 839 GPPGAPGAPGPPGPIGNVAGPKAGARGSGAGPPGATGFFGAAGRVGPPGSGNAGPPGPP 898  
 QY 562 GPAGP---AGERGEGSPGSGFQGLPGAPGAGKSGRGGVGGDLGAPGSGGAPG 618  
 Db 899 GPAGKGGKGRGETGAPRGEVGPVPPGPPGAGEKSGFADGAPGAPGTPGPGIAGQR 958  
 QY 619 GPTGLPSPGSGRGGSGRFFGADGVAGPKGAPGAGERSGGPAGP---KXSGEAGRPGEA 675  
 Db 959 GVVGLPQGRGERFFGLPSPGSEPKQSGASGERGPPGPPGGLAGPPGSGSGREGAP 1018  
 QY 676 GLPAGKGLTSGSPGPDGKTPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEFG 735  
 Db 1019 GAEGSPGSGRGGSGRGGTGTGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 1078  
 QY 736 GXAGERGVPPGAVGAPGAGKDG-----AGAGGPPG-----AGPA 771  
 Db 1079 GPAGARGPAGPPGPDGKGTGEGQDRIKGRHSGSLGPPGPPGSGEGSPGASGSPA 1138  
 QY 772 GERGGQGPAGSP---CFQGLPGAPGAGKSGRGGVGGDLGAPGSGGAPG 820  
 Db 1139 GPRGPPGSGAGAPKDKGLNGLPGIPGPPGRGTGADGVVGPVPPGPPGPPG 1190

## RESULT 8

US-10-357-851-1  
 ; Sequence 1, Application US/10357851  
 ; Publication No. US20040151731A1  
 ; GENERAL INFORMATION:



```

Db 659 GEQGVPGDLGAPGSGARGGFPGERGVQGGPPGAPRGANGAPGNDGAKGADGAPGAP 718
QY 424 GERGGPSRGFFGADGAVGKGPAGERSGPPAGPKGSPGE-----AGRGEAGLP 474
Db 719 GSGGAPGLOCMGGERGAAGLPKGDGADGPKGADGSPGKGVRLGTGPIGPPGAPGAP 778
QY 475 GAK-----GLTSGPSGPPDGKTGPPGAGODGRPPGPPGARGQ-----515
Db 779 GDKSGSPGSPAGPTGARGAPGDRGPPGPPGAGFAGPGADGQSGAKGEPGDAGAKGDA 838
QY 516-----AGVMGFGPKGAGEPKAGERGVPGPGR-----VGPAGKGEAGAAQGP 561
Db 839 GPPGAPGAPGPPPIGNVAPAKAGSAGPPGATGFFGAAGRVGPPGPPSGNAGPPGP 898
QY 562 GPAGP-----AGERGQGPAGSPGFOGLPGPAGPPGAGKPGGQGVPGDLGAPGSPGAPG 618
Db 899 GPAGKEGGKPRGETGAPRGVEVGPAGGKSGFADGAGAGCTGPGQGIAGOR 958
QY 619 GPTGLPGPPGRCGSGRFFGADGAVGKGPAGRGSPGAPG-----KGSPEAGRPGEA 675
Db 959 GVVGLPGORGERGFPGLPSPGEPKQGPFGSASGERGPPGPPGLAGPPGCSGREGAP 1018
QY 676 GLPGAKGLTSGPSGPPDGKTGPPGAGODGRPPGPPGARGQGVPGDLGAPGSPGAPG 735
Db 1019 GAEGSPGRDGSFAGKGDGRTGAPGPPGAPGAPGAPGVPAGKSGDRGETGAPGAPGV 1078
QY 736 KAGBRGVGPPGAVGAPGAKDGE-----AGAQGPPG-----AGPA 771
Db 1079 GPAGARGPAGPQGPREDGKGTGEQDRIKGRHGFGLQGGPPGPPGSGEOPSGASGPA 1138
QY 772 GERGQGPAGSP---GFOGLPGPAGPGEACKPGEQVPGDLGAPGSPGAPG 820
Db 1139 GPRGPPGAGAPGKDLNGLPGLPPIGPPGPRGTGDAGVPVPPGPPGPPG 1190

RESULT 10
US-10-788-792-150
; Sequence 150, Application US/10788792
; Publication No. US20040191819A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: US/10/788,792
; PRIOR FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 150
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-788-792-150

```

```

Query Match 59.6%; Score 2765; DB 17; Length 1464;
Best Local Similarity 55.4%; Pred. No. 4.8e-131;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGEPGTGLPFPGERGGSGRFFGADGAVGKGPAGERSGPPAGPKGSPGEA---57
Db 179 GPMGSPGPRGLPFPFAGPQGFQGPGEPCGASGPMGPPGPPGKNGDDGEAGKP 238
QY 58 GRPEAGLP---GAKGLTSGPSGP-----DGKTGPPAGODGRPPGPPGAR 105
Db 239 GRPERPPFPQARGLPGTAGLPKMKHGRFGSLDGAAGDAGPAGKPPGSPGENGAP 298
QY 106 QGAVMGFPKGAAGSPGKAGER-----GVPPGPAVGPA-----GKDXGAGAQ 150
Db 299 GQMGPRGLPGERGPPGAPGARGNDGATCAAGPPGPTGAPGPPGPPGAVGAKGAGPQ 358

```

```

QY 151 GPPGAPGAGERGEQGP-----AGSPGQGLPGPA 180
Db 359 GPRSGSPQVRGEPFPPGAGAGPAGNPGADGQGAAGANGACIAGAPFPGARGPS 418
QY 181 GPPGEAGKPEBQGVPGDLGAP-----GPSGPA-----GEPGPT 213
Db 419 GPQGGPPGPKNSCEPAGSPGKGTAKGEPGPVGVQPPGAGEGKRGARCEPPT 478
QY 214 GLPGPGERGGSGRFFGADGAVGKGPAGRGSPGAPGKSGPGEAGRPGEAGLPKAK 273
Db 479 GLPGPGERGGSGRFFGADGAVGKGPAGRGSPGAPGKSGPGEAGRPGEAGLPKAK 538
QY 274 GLTSGPSGPPDGKTGPPGAPGQDGRPPGPPGARGQAGVMGFPKPKGAAGEPKAGER 333
Db 539 GLTSGPSGPPDGKTGPPGAPGQDGRPPGPPGARGQAGVMGFPKPKGAAGEPKAGER 598
QY 334 GVPGPPGAVGPPAGKDEAGAAQGPAGPAGRGEGGPPGPPGQGLPGPAGPPGEAGKP 393
Db 599 GVPGPPGAVGPPAGKDEAGAAQGPAGPAGRGEGGPPGPPGQGLPGPAGPPGEAGKP 658
QY 394 GEQGVPGDLGAPGSPGAGE-----PGP-----TCLPGPP 423
Db 659 GEQGVPGDLGAPGSPGAGERGFPGERGVQGGPPGAPRGANGAPGNDGAKGADGAPAP 718
QY 424 GERGGPSRGFFGADGAVGKGPAGERSGPPAGPKGSPGE-----AGRPEAGLP 474
Db 719 GSGGAPGLQCMFGERGAAGLPKPKGDRDAGPKGADGSPGKGVRLGTGPIGPPGAPAG 778
QY 475 GAK-----GLTSGPSGPPDGKTGPPGAGODGRPPGPPGARGQ-----515
Db 779 GDXSGSPGAPGTGARGAPGDRGEPGPPGAPGPPGADGQGAAGEPGDDAGAKGDA 838
QY 516-----AGVMGFPKGAAGEPKAGERGVPPGPA-----VGPAGKDEAGAAQGP 561
Db 839 GPPGAPGAPGPPPIGNVAPAKAGSAGPPGATGFFGAAGRVGPPGPPSGNAGPPGP 898
QY 562 GPAGP-----AGERGQGPAGSPGFOGLPGPAGPPGAGKPGEQVPGDLGAPGSPGAPG 618
Db 899 GPAGKEGGKPRGETGAPRGVEVGPAGGKSGFADGAGAGCTGPGQGIAGOR 958
QY 619 GPTGLPGPPGRCGSGRFFGADGAVGKGPAGERSGPPAGPKGSPGEA---57
Db 1019 GAEGSPGRDGSFAGKGDGRTGAPGPPGAPGAPGAPGVPAGKSGDRGETGAPGAPGV 1078
QY 736 KAGBRGVGPPGAVGAPGAKDGE-----AGAQGPPG-----AGPA 771
Db 1079 GPAGARGPAGPQGPREDGKGTGEQDRIKGRHGFGLQGGPPGPPGSGEOPSGASGPA 1138
QY 772 GERGQGPAGSP---GFOGLPGPAGPGEACKPGEQVPGDLGAPGSPGAPG 820
Db 1139 GPRGPPGAGAPGKDLNGLPGLPPIGPPGPRGTGDAGVPVPPGPPGPPG 1190

```

```

RESULT 11
US-10-468-091-25
; Sequence 25, Application US/10468091
; Publication No. US20040157329A1
; GENERAL INFORMATION:
; APPLICANT: ADP Pharmaceutical Pty Limited
; APPLICANT: The University of Sydney
; TITLE OF INVENTION: Matrix gene expression in chondrogenesis
; FILE REFERENCE: 500311
; CURRENT APPLICATION NUMBER: US/10/468,091
; CURRENT FILING DATE: 2003-08-13
; PRIOR APPLICATION NUMBER: AU PR3116
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25

```

; LENGTH: 1461  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-468-091-25

Query Match 59.5%; Score 2761; DB 16; Length 1461;  
Best Local Similarity 55.3%; Pred. No. 7.5e-131;  
Matches 560; Conservative 31; Mismatches 229; Indels 192; Gaps 19;

```
QY 1 GPPGEPGTLPGPPGSGSGFPCADQVAGPKPAGERSGPGAGPKGSPGEA--- 57
Db 176 GPMGSGRGLPGPPGAPPGPGFPGCEPEPGASGPMGRPPGPKNGDDGEAGKP 235
QY 58 GRPGEAGL---GAKGLTSGSGSPG---DGKTGPPGAGQDGRPPGPPGAR 105
Db 236 GRPGERGPPGQAGRLPGTAGLFGMKHGRGSLDGAKGDAGAGPKGEPGPGENGAP 295
QY 106 GQAGVMGPPGKGAAGEPKKAGER-----GVPPFGAVGPA-----GKDGEAGQ 150
Db 296 GQMGRGLPGRGPPGAPGAPGAGRGNDGATGAAGPPGTPGAGPPGPFAGVAGKGEAGPQ 355
QY 151 GPPGAPGAGRGQGP-----GPSGA-----GPGPT 213
Db 416 GPQPGGPPGPKNSGEPGAFSGKDTGAKGEPGVGVQPPGAGBEGKRGARGEFPT 475
QY 214 GLPDPGRRGPGSGRFGAGVAGPKPAGERSGPGAGPKGSPGAGRPGEAGLPKAX 273
Db 476 GLPDPGRRGPGSGRFGAGVAGPKPAGERSGPGAGPKGSPGAGRPGEAGLPKAX 535
QY 274 GLTSGSGSPGDKTGTGPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEPKKAGER 333
Db 536 GLTSGSGSPGDKTGTGPPGAGQDGRPPGPPGARGQAGVMGPPGKGAAGEPKKAGER 595
QY 334 GVPDPGAVGPKDCEAGAGPPGPPGAGPAGEREGPAGSPGQGLPGRAGPGEAGKP 393
Db 596 GVPDPGAVGPKDCEAGAGPPGPPGAGPAGEREGPAGSPGQGLPGRAGPGEAGKP 655
QY 394 GEGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
Db 656 GEGVPGDLGAPGSPGAGERGFFGCGVQPPGPPGPRGANGAGNDGAKGDAGAPG 715
QY 424 GERGPGSRGPPGAGVAGPKPAGERSGPGAGPKGSPG-----AGRPGEAGLP 474
Db 716 GSQGAPGLQGMGERGAAAGLPKPKDGRDAGPKGADGSGFKDGVRLTGTPIGPPGAP 775
QY 475 GAK-----GLTSGSGSPGDKTGTGPPGAGQDGRPPGPPGARGQ----- 515
Db 776 GDXGSGSPGAPGTGARGAPDRGEPGPPGAGPPGADGQPGAKGEGFDGAKGDA 835
QY 516 -----AGVMGFPKGAAGEPKKAGERGVPPGPA-----VGPAGKDGAGAGPP 561
Db 836 GPPGAPGAPGPPGPIGNVAPCAKAGSAGPPGATGFFGAAGRVPGPSGNAGPPGP 895
QY 562 GPAGP-----AGEREGPAGSPFQGLPGPAGPGEAGKPGCQVPGDLGAPGSPGAGEP 618
Db 896 GPAGKEGKGKRGETGPAGRPGEVPPGPPGAGEKSGPDADGAPAGPCTPGPGIAGQR 955
QY 619 GPTGLPPPGRRGSGSRGFPAGDGVAGPKPAGERSGPPGAPG---XGSPGEAGRPGEA 675
Db 956 GVVGLPGQRRGFFCLPGPSGPKQKFSGASGERGPPGPMGPPGLAGPPGESGEGAP 1015
QY 676 GLPAGKLTGSGSGSPGDKTGTGPPGAGQDGRPPGPPGARGQAGVMGFPKGAAGEP 735
Db 1016 GAEGSPGRDGSFCAKDRGETGAPGPPGAPGAPGFPVGPAGKSGDRGETGAPAGPV 1075
QY 736 GKAGRGVPPGAVGVPAGKDE-----AGAQGGPP-----AGPA 771
Db 1076 GVPKAGPAGPQGRDKGETGEQDGRGIKHGRGSLGQGPFGPPGSGFGEQGPSASCAPA 1135
```

```
QY 772 GEREGQGPAGSP---GFGQLPGPAGPPGEGAGKPKGQGVPGDLGAGPSPGAPG 820
Db 1136 GPRGPPGAGAPKDGKGLNGLPGPIGPPGRGRTGDAGVPVGPGRPPGPPG 1187
```

## RESULT 12

US-10-291-265-243  
; Sequence 243, Application US/10291265  
; Publication No. US20030232054A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; APPLICANT: Tang et al.  
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides  
; FILE REFERENCE: 21272-017 (785)  
; CURRENT APPLICATION NUMBER: US/10/291,265  
; CURRENT FILING DATE: 2000-01-25  
; PRIOR APPLICATION NUMBER: 09/491,404  
; PRIOR FILING DATE: 2000-01-25  
; PRIOR APPLICATION NUMBER: 09/617,746  
; PRIOR FILING DATE: 2000-07-17  
; PRIOR APPLICATION NUMBER: 09/631,451  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 09/633,870  
; PRIOR FILING DATE: 2000-09-15  
; NUMBER OF SEQ ID NOS: 944  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 243  
; LENGTH: 1464  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-291-265-243

Query Match 59.5%; Score 2761; DB 14; Length 1464;

Best Local Similarity 55.3%; Pred. No. 7.5e-131;  
Matches 560; Conservative 31; Mismatches 229; Indels 192; Gaps 19;

```
QY 1 GPPGEPGTLPGPPGSGSGFPCADQVAGPKPAGERSGPGAGPKGSPGEA--- 57
Db 179 GPMGSGRGLPGPPGAPPGPGFPGCEPEPGASGPMGRPPGPKNGDDGEAGKP 238
QY 58 GRPGEAGL---GAKGLTSGSGSPG---DGKTGPPGAGQDGRPPGPPGAR 105
Db 239 GRPGERGPPGQAGRLPGTAGLFGMKHGRGSLDGAKGDAGAPGPKGEPGPGENGAP 298
QY 106 GQAGVMGFPKGAAGEPKKAGER-----GVPPFGAVGPA-----GKDGEAGQ 150
Db 299 GQMGPRGLPGERGPPGAPGAPGAGRGNDGATGAAGPPGTPGAGPPGPGAVGAKGAGPQ 358
QY 151 GPPGAPGAGRGQGP-----GPSGA-----AGSPGQGLPGPA 180
Db 359 GPRGSEGPQVGRGEPGPPGAGAAAGPAGNPGADGQPGAKGANGAPGIAGAPGPPGARGPS 418
QY 181 GPPGEGAKTGEQGVPGDLGAP-----GPSGA-----GPPGPT 213
Db 419 GPQPGGPPGPKNSGEPGAPGSKGTGAKGEPGVGVQPPGPPGAGEBKRGARGEFPT 478
QY 214 GLPDPGRRGPGSGRFGAGVAGPKPAGERSGPPGAGPKGSPGAGRPGEAGLPKAX 273
Db 479 GLPDPGRRGPGSGRFGAGVAGPKPAGERSGPPGAGPKGSPGAGRPGEAGLPKAX 538
QY 274 GLTSGSGSPGDKTGTGPPGAGQDGRPPGPPGARGQAGVMGFPKGAAGEPKKAGER 333
Db 539 GLTSGSGSPGDKTGTGPPGAGQDGRPPGPPGARGQAGVMGFPKGAAGEPKKAGER 598
QY 334 GVPDPGAVGPKDCEAGAGPPGPPGAGPAGERSGPPGAPG---XGSPGEAGRPGEA 675
Db 599 GVPDPGAVGPKDCEAGAGPPGPPGAGPAGERSGPPGAPG---XGSPGEAGRPGEA 658
QY 394 GEGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
Db 659 GEGVPGDLGAPGSPGAGERGFFGCGVQPPGPPGPRGANGAGNDGAKGDAGAPG 718
QY 424 GERGPGSRGFPAGDGVAGPKPAGERSGPPGAGPKGSPG-----AGRPGEAGLP 474
```







Db 359 GPRGSEPGQVGRGEPGPGPAGAGPAGNPGADQPGAKGANGAPGIAGAPGFFGARGFS 418  
QY 181 GPPGSAKPGEOGVPGLGAP-----GPSGPA-----GEPGPT 213  
Db 419 GQGGPGPPGPKNSGEPGAPGSGDKTGAKGEPGVGVGGPPGPAEGEGKARGEPGPT 478  
QY 214 GLPGPPGERGGSGFFPGADGAVGPKGPAGRGSPGPAKGSGEAGRPCEAGLPKAK 273  
Db 479 GLPGPPGERGGSGRGGFFGADGAVGPKGPAGRGSPGPAKGSGEAGRPCEAGLPKAK 538  
QY 274 GLTSGSGSPGPKTKGPPGAGQDGRPGPPGPGARGOAGVMGPPGKGAAGEBCKAGER 333  
Db 539 GLTSGSGSPGPKTKGPPGAGQDGRPGPPGPGARGOAGVMGPPGKGAAGEBCKAGER 598  
QY 334 GVPGGPAGVGPAGKDEAGAGQPPGAGPAGRGSPGPAKGSGEAGRPCEAGKP 393  
Db 599 GVEGPPGAVGPAGKDEAGAGQPPGAGPAGRGSPGPAKGSGEAGRPCEAGKP 658  
QY 394 GEQGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423  
Db 659 GEQGVPGDLGAPGSPGAGERGFPGERGVQGGPPGPRGANGAPGNDGAKGDAGAPAP 718  
QY 424 GERGGPSRGGFADGVAGPKGPAGRGSPGPAKGSGEAGRPCEAGKP 474  
Db 719 GSQAGPGLQMGEGRGAAGLPKPKGDRGDAGPKGADGSPGDKGVRLTGPIGPPGAP 778  
QY 475 GAK-----GLTSGSGSPGPKTKGPPGAGQDGRPGPPGARGO-----515  
Db 779 GSKGSGSPGAPGPTGARGAGDRGEPGPPGAGPAGRGSPGPAKGSGEAGRPCEAGKP 838  
QY 516 -----AGVMGPPGKGAAGEBCKAGERGVPPGPA-----VGPAGKDEAGAGQPP 561  
Db 839 GPPGAPAGPFPPIGNVGAAGKARGASAGPPGATGPPGAAGRVGPPGSGNAGPPGPP 898  
QY 562 GPAGP---AGERGQSPGSPGQGLPGPAGPAGRGSPGPAKGSGEAGRPCEAGKP 618  
Db 899 GPAGKEGKGRGETGPRGPEVGPFGPPGPAKGSGEAGRPCEAGKP 958  
QY 619 GFTGLPGPPGERGGSGFTGADGAVGPKGPAGRGSPGPAKGSGEAGRPCEAGKP 675  
Db 959 GVUGLPGORGERFFGLPGSGEPGKQSPGASGERGPPGPMGPPGLAGPPGSGREGAP 1018  
QY 676 GLPGAKGLTSGSPGPKTKGPPGAGQDGRPGPPGARGOAGVMGPPGKGAAGEP 735  
Db 1019 AAGSGSPGRDGSFGAKGDRGETGPPGAPGAPGAPGVPVGPAGKSGDRGETGPPGAPGAPV 1078  
QY 736 GKAGERGVPPGPAVGPAGKDE-----AGAQPPG-----AGPA 771  
Db 1079 GPGVARGPAGPQGRGDKGETGQGDRIKGRHGFSLQGLGPPGPPGSGPSGASGPA 1138  
QY 772 GERGQGPAGSP---GFGPLPGPAGPAGRGSPGPAKGSGEAGRPCEAGKP 820  
Db 1139 GPRGPPGAGAPGKDLNGLPGIPGPGPRGTGADGVPVGPFGPPGPPG 1190

Search completed: October 18, 2004, 13:58:35  
Job time : 112.622 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:31:10 ; Search time 32.5431 Seconds  
(without alignments)  
2427.363 Million cell updates/sec

Title: US-10-658-989A-3  
Perfect score: 4640  
Sequence: 1 GPGGPGTGLGPPGCGG.....GEGVPGDLGAPGSPGAGG 821  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:.\*  
1: PIR1:.\*  
2: PIR2:.\*  
3: PIR3:.\*  
4: PIR4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2755	59.4	1464	1 CGHU1S	collagen alpha 1(I)
2	2732.5	58.9	1453	2 S21626	collagen alpha 1(I)
3	2679	57.7	1042	1 CGHU1S	collagen alpha 1(I)
4	2477	53.4	1418	2 T45467	collagen alpha 1(I)
5	2476	53.4	1487	1 CGHU6C	collagen alpha 1(I)
6	2460	53.0	1419	2 A41182	collagen alpha 1(I)
7	2460	53.0	1487	2 B41182	collagen alpha 1(I)
8	2401	51.7	1486	1 B40333	collagen alpha 1(I)
9	2389	51.5	1492	2 A40333	collagen alpha 1(I)
10	2379	51.3	1049	1 CGS07S	collagen alpha 1(I)
11	2377.5	51.2	1464	2 S59856	collagen alpha 1(I)
12	2370	51.1	1466	1 CGHU7L	collagen alpha 1(I)
13	2326	50.1	1496	1 CGHU2V	collagen alpha 2(V)
14	2303.5	49.6	1497	2 I49607	procollagen type V
15	2300	49.6	1373	1 A43291	collagen alpha 2(I)
16	2263.5	48.8	1366	1 CGHU2S	collagen alpha 2(I)
17	2235	48.2	671	1 CGHT1S	collagen alpha 1(I)
18	2229.5	48.0	1838	1 CGHU1V	collagen alpha 1(V)
19	2208.5	47.6	1843	2 S18803	collagen alpha 1(V)
20	2193.5	47.3	779	1 CGB01S	collagen alpha 1(I)
21	2169	46.7	1806	1 CGHU1E	collagen alpha 1(X)
22	2161.5	46.6	886	2 I50694	collagen alpha 2(I)
23	2136.5	46.0	1414	1 S23809	collagen alpha 2(I)
24	2074.5	44.7	1027	2 S28774	collagen alpha 1(V)
25	2071	44.6	2944	2 A54849	collagen alpha 1(V)
26	2060.5	44.4	1546	1 CGHU2E	collagen alpha 2(X)
27	2058	44.4	1691	1 S22917	collagen alpha 5(I)
28	2046	44.1	1690	1 CGHU1B	collagen alpha 4(I)
29	2036	43.9	1549	2 I48103	type VII collagen

## RESULT 1

## CGHU1S

collagen alpha 1(I) chain precursor - human

N;Alternate names: procollagen alpha 1(I) chain

C;Species: Homo sapiens (man)

C;Date: 12-Aug-1981 #sequence revision 04-Oct-1996 #text change 09-Jul-2004

C;Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S1

5269; A29439; I53466; A02852; I37247

R;D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J

Gene 67, 105-115, 1988

A;Title: Complete nucleotide sequence of the region encompassing the first twenty-five

A;Reference number: I60114; MUID:88329734; PMID:2843432

A;Accession: I60114

A;Status: translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-369, 'L', 371-589 <DAL>

A;Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNI

R;Iromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Proc

Biochem. J. 253, 919-922, 1988

A;Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human

A;Reference number: S01143; MUID:89025644; PMID:3178743

A;Accession: S01143

A;Molecule type: mRNA

A;Residues: 1-472 <TRO>

A;Cross-references: EMBL:X07884; NID:G30015; PID:CAA30731.1; PID:G30016; GB:M36546; NI

A;Note: Submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C

Nature 310, 337-340, 1984

A;Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation o

A;Reference number: A93335; MUID:84270697; PMID:6462220

A;Accession: A93335

A;Molecule type: DNA

A;Residues: 1-58, 'Q', 60-181 <CHU>

A;Cross-references: EMBL:X00820; NID:G35657; PID:CAA2394.1; PID:G35658

R;Roscow, C.M.S.; Vergeer, W.P.; du Picoy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W

J. Biol. Chem. 262, 15151-15157, 1987

A;Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene en

A;Reference number: I55254; MUID:88033098; PMID:2822714

A;Accession: I55254

A;Status: translation not shown; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-45 <ROS>

A;Cross-references: GB:J02829; NID:G180387; PID:AAA51993.1; PID:G180388

R;Bornstein, P.; McKay, J.; Morishima, J.K.; Devaryalu, S.; Gelinas, R.E.

Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987

A;Title: Regulatory elements in the first intron contribute to transcriptional control

A;Reference number: A39943; MUID:88097389; PMID:3480516

A;Accession: A39943

A;Molecule type: DNA

A;Residues: 1-34 <BOR>

A;Cross-references: GB:J03559; NID:G180876; PID:AAA52052.1; PID:G553238

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

- J. Biol. Chem. 260, 2315-2320, 1985  
A;Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s  
A;Reference number: 155237; MUID:85130970; PMID:2857713  
A;Accession: 155237  
A;Molecule type: mRNA  
A;Status: translation not shown; translated from GB/EMBL/DBJ  
A;Residues: 1-34 <CH2>  
A;Cross-references: GB:M10627; NID:G180383; PIDN:AAAS1992.1; PID:9553226  
R;Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollis  
J. Biol. Chem. 265, 6312-6317, 1990  
A;Title: In vivo and in vitro noncovalent association of excised alpha 1(I) amino-terminal  
some type VII.  
A;Reference number: A35233; MUID:90202908; PMID:2318855  
A;Accession: A35233  
A;Molecule type: protein  
A;Residues: 33-52 <WR>  
A;Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved  
R;Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.  
EMBO J. 8, 1705-1710, 1989  
A;Title: A base substitution in the exon of a collagen gene causes alternative splicing  
A;Reference number: S09400; MUID:89356643; PMID:2767050  
A;Accession: S09400  
A;Molecule type: mRNA  
A;Residues: 156-183 <WEI>  
R;Click, E.M.; Bornstein, P.  
Biochemistry 9, 4699-4706, 1970  
A;Title: Isolation and characterization of the cyanogen bromide peptides from the alpha  
A;Reference number: A30567; MUID:71038625; PMID:5529814  
A;Contents: CNB\*0-1, CNB\*2, CNB\*4, CNB\*5  
A;Accession: B90567  
A;Molecule type: protein  
A;Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'  
A;Note: evidence for 170-allysine  
R;Baetge, B.; Notbald, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, R.  
Eur. J. Biochem. 192, 153-159, 1990  
A;Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle  
A;Reference number: S11372; MUID:90382436; PMID:2169412  
A;Accession: S11372  
A;Molecule type: protein  
A;Residues: 173-187; 274-287, 'P', 289 <BAE>  
A;Note: sequence of collagen alpha 1(S) (I) isolated from bone after pepsin digestion  
R;Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez  
J. Biol. Chem. 266, 21827-21832, 1991  
A;Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain  
cooperative melting of intact type I collagen.  
A;Reference number: 155342; MUID:92042092; PMID:1718984  
A;Accession: 155342  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 258-268; 1347-1357 <DEA>  
A;Cross-references: GB:S67495; NID:G239007; PIDN:AB20350.1; PID:G239008  
A;Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report  
R;Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.  
J. Biol. Chem. 245, 5042-5048, 1970  
A;Title: Comparative study of glycopeptides derived from selected vertebrate collagens.  
A;Reference number: A92069; MUID:71001508; PMID:4319110  
A;Accession: A92069  
A;Molecule type: protein  
A;Residues: 263-268 <MOP>  
A;Experimental source: skin  
A;Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxyllysine  
R;Labhard, M.E.; Hollister, D.W.  
Matrix 10, 124-130, 1990  
A;Title: Segmental amplification of the entire helical and telopeptide regions of the cd  
A;Reference number: S15989; MUID:90326017; PMID:2374517  
A;Accession: S15989  
A;Molecule type: mRNA  
A;Residues: 281-302; 402-420; 823-843; 925-944; 1026-1045; 1143-1162 <LAB>  
R;Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N.  
Connect. Tissue Res. 29, 1-11, 1993  
A;Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of  
A;Reference number: 152905; MUID:93339042; PMID:8339541  
A;Accession: 152905  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 342-352, 'C', 354-359 <W12>  
A;Cross-references: GB:S64717; NID:G408195; PIDN:AB27677.1; PID:G408196  
A;Note: mutant sequence from patient with osteogenesis imperfecta  
R;Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Ekenberry, E.F.; Prockop, D.J.  
Biochemistry 22, 5213-5223, 1983  
A;Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha  
A;Reference number: A90476; MUID:84080385; PMID:6689127  
A;Accession: A90476  
A;Molecule type: mRNA  
A;Residues: 425-1250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>  
A;Cross-references: GB:K01228; NID:G180391; PIDN:AAAS1995.1; PID:G180392  
A;Note: sequence partially completed for missing nucleotides by A29439  
R;Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.  
J. Biol. Chem. 260, 691-694, 1985  
A;Title: Multixon deletion in an osteogenesis imperfecta variant with increased type II  
A;Reference number: A22161; MUID:85104934; PMID:2381843  
A;Accession: A22161  
A;Molecule type: DNA  
A;Residues: 472-594, 'R', 596-607 <CH3>  
A;Cross-references: GB:K03178; GB:K03179; NID:G179612; NID:G179613; PIDN:AAAS1847.1; PID  
A;Note: the authors translated the codon CGT for residue 595 as Pro  
R;Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.  
Am. J. Hum. Genet. 46, 1034-1040, 1990  
A;Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained  
A;Reference number: A35336; MUID:90252792; PMID:2339700  
A;Accession: A35336  
A;Molecule type: mRNA  
A;Residues: 710-720, 'E', 722-737, 'E', 739-745 <WAL>  
A;Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu  
R;Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes  
Hum. Mol. Genet. 3, 2201-2206, 1994  
A;Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the c  
A;Reference number: 154365; MUID:95187161; PMID:7881420  
A;Accession: 154365  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 746-766, 'S', 768-781 <FOR>  
A;Cross-references: GB:L47687; NID:G1009093; PIDN:AB59576.1; PID:G1009094  
R;Chesler, S.D.; Wallis, G.A.; Byers, P.H.  
J. Biol. Chem. 268, 18218-18225, 1993  
A;Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty  
A;Reference number: A47426; MUID:93352646; PMID:8349697  
A;Accession: A47426  
A;Molecule type: mRNA  
A;Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>  
A;Cross-references: GB:S84596; NID:G407589; PIDN:AB27856.1; PID:G407590  
A;Note: sequence extracted from NCBI backbone (NCBI:136444, NCBI:136445)  
A;Note: does not represent an experimentally determined sequence but three different mut  
A;Accession: B47426  
A;Molecule type: mRNA  
A;Residues: 1179-1464 <CH4>  
A;Experimental source: normal dermal fibroblast culture  
A;Accession: C47426  
A;Molecule type: mRNA  
A;Residues: 1179-1276, 'H', 1278-1464 <CH5>  
A;Experimental source: fetal cell 86-237  
A;Accession: D47426  
A;Molecule type: mRNA  
A;Residues: 1179-1336, 1339-1464 <CH6>  
A;Experimental source: fetal cell 86-146  
A;Accession: E47426  
A;Molecule type: mRNA  
A;Residues: 1179-1387, 'R', 1389-1464 <CH7>  
A;Experimental source: fetal cell 88-251  
R;Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Nid  
J. Biol. Chem. 263, 14605-14607, 1988  
A;Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptide c  
A;Reference number: 155269; MUID:89008319; PMID:3170557  
A;Accession: 155269

A;Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1187-1194, 'C', 1196-1220 <COH>  
A:Cross-references: GB:M23213; NID:G440842; PIDN:AA859363.1; PID:G499622  
A:Note: mutant sequence from a patient with mild osteogenesis imperfecta  
R;Mackelae, J.K.; Raasina, M.; Virta, A.; Vuorio, E.  
Nucleic Acids Res. 16, 349, 1988  
A>Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

Query Match 59.4%; Score 2756; DB 1; Length 1464;  
Best Local Similarity 55.3%; Pred. No. 3.5e-130;  
Matches 560; Conservative 31; Mismatches 229; Indels 192; Gaps 19;

Qy 1 GPPGEPPTGLPGPPGGRGSGRFPFGADGAVGKPGAGERSGPGAGPKGSPGGEA--- 57  
Db 179 GPMGSGRGLPGPPAGPGPGQFGPGEGPCASGPMGPRGPPGPKNGNDGEGAKP 238  
Qy 58 GRPEAGLPG---GAKLTGSPGSGP-----DCKTGPAGQDGRGPPGPGAR 105  
Db 239 GRPGERPPGQAGRLPGTAGLPMKHRGFSGLDGAKDAGPAGPKGSPGPGENGAP 298  
Qy 106 QGAGVMGFPFGKGAAGFPKAGER-----GVPPPGAVGPA-----GKDGAGAG 150  
Db 239 QMGPRGLPGRGPRGPGAGARGNDGATGAAGPGTGPAGPPGPGAVGAKGAGAPQ 358  
Qy 151 GPPGAGPAGRGSGP-----AGSPGQGLPCPA 180  
Db 359 GPRGSEGGVGRGPPGPPGAGAGAGNPGADGQCGAKGANGAPGAGPFGGARGPS 418  
Qy 181 GPPGAGKPGQGVGDLGAP-----GPSGPA-----GEGGPT 213  
Db 419 GPGGPPGPPGKNGSGEPGAGSGKDTGAKGEPGVGQPPGAGBEKRGARGGPT 478  
Qy 214 GLPFPGRGGRGSPGADGAVGKPGAGERSGPGAGPKGSPGAGRPGEAGLPGAK 273  
Db 479 GLPFPGRGGRGSPGADGAVGKPGAGERSGPGAGPKGSPGAGRPGEAGLPGAK 538  
Qy 274 GLTSGPGRPDCKTGPAGODGRPPGARGOAGVMGPPGKGAAGEPKAGER 333  
Db 539 GLTSGPGRPDCKTGPAGODGRPPGARGOAGVMGPPGKGAAGEPKAGER 598  
Qy 334 GVPGPPGAVGPKDGEAGAGQPPGAGPAGERGEGQPGAGSPGQGLPGPAGPGEAGKP 393  
Db 599 GVPGPPGAVGPKDGEAGAGQPPGAGPAGERGEGQPGAGSPGQGLPGPAGPGEAGKP 658  
Qy 394 GEGQVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 423  
Db 659 GEGQVPGDLGAPGSGPARGERFPGERGYPGPPGAPRGANGAPNDGAKGDAGAPAP 718  
Qy 424 GERGGPGRGFPAGDVGKPGAGERSGPGAGPKGSPGE-----AGRPGEAGLP 474  
Db 719 GSGGAPGLQGMPEERGAAGLPKGDGADGPKGADGSGKDGVRGLTGP.GPQPGAP 778  
Qy 475 GAK-----GLTSGPSPGPDGKTGTPPGAGQDGRPPGPPGARGQ----- 515  
Db 779 GDKGESGSPGAPGTGARGAPGDRGEPGPPGAPGAGPPGADGQPGKAGPGDAGAKGDA 838  
Qy 516 -----AGVMGFPKGAAGERKAGERGYPGPPGA-----VGPAGKGEAGAGQPP 561  
Db 839 GPPGAPGAPGPPGPIGVGAPKAGARGAGSPGATGPPGAAGRVGPPGPGSNAGPPGPP 898  
Qy 562 GPAGP---AGERGEQGPAGSPGQGLPGPAGPFGAGKEGEGQVFGDLGAPGSPGAPGE 618  
Db 899 GPAGKGGKGRGETGTPAGRPGEVGPFGPPGAGEKSPGADGAPAGTGPQGIAGOR 958  
Qy 619 GPTGLPGPGRGGRGSPGADGAVGKPGAGERSGPPAGP---KQSPGEAGRPGEA 675  
Db 959 GVVGLPGQGRGERGFPGLPGSPGPPGQKQPSGASGERGPPGPPGLGAPPGESGREGAP 1018  
Qy 676 GLPGAKGLTSGPSPGPDGKTGTPGAGQDGRPPGPPGARGQAGVMGFPFGKGAAGEP 735  
Db 1019 GAEGRDGRGSPGAKGDRGETGAPGPPGAXGAXGAPGVPGAPKSGDRGETGAPGAPGV 1078

Qy 736 KGAGRGVPGPPGAVGVPAGKDG-----AGQGGPPG-----AGPA 771  
Db 1079 CPAGARGPAGPQPRGDKGETGEGQDRGKIGHRGFSGLQGGPPGPGSGEQPGSASGPA 1138  
Qy 772 GERGQGGPAGSP---GFGGLPGPAGPGEAGKPGEQGVGDLGAPGSPGAPG 820  
Db 1139 GPRGPPGAGAPCKDGLNGLPGPIPGPRGTGADGVPGPPGPPGPPG 1190

RESULT 2  
collagen alpha 1(I) chain precursor - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text change 09-Jul-2004  
C;Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626  
R;Li, S.W.; Khillan, J.; Prockop, D.J.  
Matrix Biol. 14, 593-595, 1994  
A>Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I  
A;Reference number: S57243  
A;Accession: S57243  
A:Molecule type: mRNA  
A:Residues: 1-1453 <LIS>  
A:Cross-references: UNIPROT:P11087; EMBL:U08020; NID:G470673; PIDN:AAA88912.1; PID:G470  
R;Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.  
Biochim. Biophys. Acta 1089, 241-243, 1991  
A>Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs  
A;Reference number: S16176; MUID:91274355; PMID:2054384  
A;Accession: S16374  
A;Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1442-1453 <MET>  
A:Cross-references: EMBL:X57981; NID:G50484; PIDN:CAA41046.1; PID:G50485  
R;French, B.T.; Lee, W.H.; Maul, G.G.  
Gene 39, 311-312, 1985  
A>Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.  
A;Reference number: A23982; MUID:86137403; PMID:3841523  
A;Accession: A23982  
A:Molecule type: mRNA  
A:Residues: 518-1128 <PRE>  
A:Cross-references: GB:M14423; NID:G192261; PIDN:AAA37333.1; PID:G192262  
R;Monson, J.M.; Friedman, J.; McCarthy, B.J.  
Mol. Cell. Biol. 2, 1362-1371, 1982  
A>Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene. Evidence for  
A;Reference number: I49559; MUID:83141374; PMID:6298597  
A;Accession: I49559  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 735-1130 <RES>  
A:Cross-references: GB:M17491; NID:G192263; PIDN:AAA37334.1; PID:G192264  
R;Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.  
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984  
A>Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads  
A;Reference number: I49557; MUID:84170331; PMID:6324198  
A;Accession: I49557  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-25 <RES>  
A:Cross-references: GB:X01688; NID:G192246; PIDN:AAA37330.1; PID:G553881  
R;Fenton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.  
Biochim. Biophys. Acta 1216, 469-474, 1993  
A>Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.  
A;Reference number: S39789; MUID:94092741; PMID:8268229  
A;Accession: S39789  
A:Molecule type: DNA  
A:Residues: 1-80, 'E', 82-105, 'D', 107-185, 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-  
R;Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.  
Mol. Cell. Biol. 14, 5950-5960, 1994  
A>Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indire  
A;Reference number: I48300; MUID:94344105; PMID:8065328  
A;Accession: I48300  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>

Query Match	58.9%; Score 2732.5; DB 2; Length 1453;
Best Local Similarity	54.7%; Pred. No. 5.1e-129;
Matches	547; Conservative 33; Mismatches 240; Indels 180; Gaps 16;
QY	1 GPPGPGTGLGPPGE-----RCGPGSGPPGADGVAGPKGPAGERSGPPA 48
Db	180 GPPGAPGCGGPGPPGPPGPPGPGSGGPMRPGPPGPGKNGDDGEAGKGRPGERPPGPQ 239
QY	49 GPKSGPGAGRP-----GEAGLPCAKGLTCSPGSPGDPCKTCTPPCPAQD 93
Db	240 GARGLEGTAGLPGMKHRRGSLDGAKGADGAPAGKGPSPGCGAGPQMGPRGLPGER 299
QY	94 GRCPGPPGPPGARGGQAVMGPPGPKGAAGE-----PKKAGERGVPGPCAVGPAKGCGEA 147
Db	300 GRPGPGGTAGARGNDGAVGAAGPPGTPGTGPPGFCAGVAKGEAGPQARGSGEPQVR 359
QY	148 GQAQPPGPPGAPAGERGQQP-----AGSPFOCLPQAPPPGPPCEAGKPGEQ 192
Db	360 CEPGPPGPAGAAGFAGNPGADGQGAKGANGAPGIAGA-PGFCARGPGSQPGSPGPGPK 419
QY	193 GVPDGLGAP-----GPGGPA-----GRPPTGLPCTPPEGRGP 225
Db	420 GNSGEPAGPNKGTGAKGFPFATGVQGGPPGABEGKRGARGEPGSGLTPGPPGERGP 479
QY	226 GSRFPFGADGVAGPKGAPAGERGSPGAPGPKGSPGAGRCEAGLPGAKGLTSGSGSPGD 285
Db	480 GSRFPFGADGVAGPKGSPGSRGAPGAPGPKGSPGAGRCEAGLPGAKGLTSGSGSPGD 539
QY	286 KTGTPGPGAGQDGRPPGPPGARGQAVMGPPGPKGAAGSPGKAGERGVPPGPAVGPA 345
Db	540 KTGTPGPGAGQDGRPPGPPGARGQAVMGPPGPKGTAGEPGKAGERGLDPPGPAVGPA 599
QY	346 GKDCGAGAQGPPGPPGAPGABERGQGPAGSPGQGLPGPAGPGEAGKPGEGVPGDGLGAP 405
Db	600 GKDCGAGAQGAPGAPGABERGQGPAGSPGQGLPGPAGPGEAGKPGEGVPGDGLGAP 659
QY	406 GPSGPAQE-----PGP-----TGLFPPGPPGGRGSGSGFP 435
Db	660 GPSGARGERFPGERGVQGGPPGAPGRNGNAPGNDGAKGDTGAPGPGSQAPGLQMP 719
QY	436 GADGVAGPKPAGERSPPGAPKSGPGE-----AGRPBGALPGAKLGTGSPGP 486
Db	720 GERGAALPGPKGRGDAGPKGADGSPGKDGARGTGTGIPGPPGAGAPGDKGEAGSPGPP 779
QY	487 GPD-----GKTGPPGAPQDGRPPGPPGARGQ-----AGVMGFP 522
Db	780 GPTGARGAPDGRGEAGPPGAPGAPGADGQPGAKGEPGDTGVKGDAGPPGAPGAPPP 839
QY	523 GPKGAAGEPKKAGERGVPPGGA-----VGPAGKDGAGACQAGPPGAPG-----AGER 570
Db	840 GPTGNVGAPGPKPRGAAGPPGATGFPGAAGRVPGPSPGNAGPPGPPGVPKGEGKGKPR 899
QY	571 GEGPAGSPGFOGLPGPAGFPGEAGKPGEQVPGDLGAPGSGPAGPFGTGLPGPPGPR 630
Db	900 GETGPAGRPGEVPPGPPFAGKSGSPGADGAPGSPGTFPGQITAGQGVVGLPGQGRG 959
QY	631 GPGSGSPFGADGVAGPKPAGERSPPGAPG-----KSPGEAGRPBAGLPGAKGLTGSPP 687
Db	960 GFFGLPGSGEPKQGPSGSSGERGPPGPMGPPGLAGPSPGREGSPGAEGSPGSDGAP 1019
QY	688 GSPGCPDKTGPFGPAGQDGRPPGPPGARGQAVMGFPFPGKGAAGPPGKAGERGVGPR- 746



A:Reference number: A38513; MUID:91184811; PMID:2081599  
A:Accession: A38513  
A:Molecule type: DNA  
A:Cross-references: <RNA>  
A:Residues: 1-103 <RYA>  
R:Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.  
Nucleic Acids Res. 17, 9473, 1989  
A:Title: Nucleotide sequence of the full length cDNA encoding for human type II procollagen  
A:Reference number: S06715; MUID:90067946; PMID:2587267  
A:Accession: S06715  
A:Molecule type: mRNA  
A:Residues: 1-28, 'R', '99-1487 <SU2>  
A:Cross-references: EMBL:X16468; NID:929515; PIDN:CAA34488.1; PID:g29516  
A:Note: alternative splice form 1  
R:Viikula, M.; Mesander, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.  
Biochem. J. 285, 287-294, 1992  
A:Title: Structural analysis of the regulatory elements of the type-II procollagen gene.  
A:Reference number: S24270; MUID:92344585; PMID:1637314  
A:Accession: S24270  
A>Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-28 <VIX>  
A:Cross-references: EMBL:X58709; GB:S40537; NID:G35659  
A:Note: this translation is not annotated in GenBank entry HSPROCOB1, release 111.0  
R:Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.  
Gene 44, 11-16, 1986  
A:Title: Promoter region of the human pro-alpha-1-(II)-collagen gene.  
A:Reference number: A24828; MUID:87031574; PMID:3021582  
A:Accession: A24828  
A:Molecule type: DNA  
A:Residues: 1-8, 'T', '10-28 <NUN>  
A:Cross-references: GB:M25698; NID:G180872; PIDN:AAAS2051.1; PID:g553237  
R:Baldwin, C.T.; Reginaldo, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.  
Biochem. J. 262, 521-528, 1989  
A:Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(II)  
A:Reference number: S06496; MUID:90026318; PMID:2803268  
A:Accession: S06496  
A:Molecule type: mRNA  
A:Residues: 7-28, 'R', '99-157, 'P', '159-440, 'G', '442-456, 'E', '458-640, 'A', '642-831, 'PA', '834, 'F', 'F'  
A:Cross-references: EMBL:X16711; NID:G30040; PIDN:CAA34683.1; PID:g30041  
R:Ryan, M.C.; Sandell, L.J.  
J. Biol. Chem. 265, 10334-10339, 1990  
A:Title: Differential expression of a cysteine-rich domain in the amino-terminal propept  
A:Reference number: A35428; MUID:90285153; PMID:2355003  
A:Accession: A35428  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 27-81, 'L', '83-103 <RYA2>  
A:Note: alternative splice form 2; splicing appears to be under developmental regulation  
R:Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.  
Genomics 4, 438-441, 1989  
A:Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf  
A:Reference number: A30147; MUID:89233138; PMID:2714801  
A:Accession: A30147  
A:Molecule type: DNA  
A:Residues: 104-157, 'P', '159-236 <SUM>  
A:Cross-references: GB:J03065; GB:M23660; GB:M25655; GB:M25656; GB:M25730; GB:M32168; GE  
R:Ala-Kokko, L.; Baldwin, C.T.; Moskowicz, R.W.; Prockop, D.J.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990  
A:Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pri  
A:Reference number: A94227; MUID:90370826; PMID:1975693  
A:Accession: A33116  
A:Molecule type: DNA  
A:Residues: 171-172, 'C', '174-175 <ALA>  
A:Note: mutant sequence from a family with primary generalized osteoarthritis  
R:Diab, M.; Wu, J.J.; Eyre, D.R.  
Biochem. J. 314, 327-332, 1996  
A:Title: Collagen type IX from human cartilage: a structural profile of intermolecular c  
A:Reference number: S64673; MUID:96195147; PMID:8660302  
A:Accession: S64674  
A:Molecule type: protein  
A:Residues: 188-189, 'X', '191-195; 1224-1230, 'X', '1232-1236 <DIA>

R;Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage, Eur. J. Biochem. 234, 125-131, 1995

A>Title: Immunohistochemical and biochemical analyses of 2000-25000-year-old fossil car

A;Reference number: S63514; MUID:96036730; PMID:8529631

A;Accession: S63514

A;Molecule type: protein

A;Residues: 243-261;575-590;756-763,'X',765-779 <FRA>

R;Tiller, G.E.; Weis, M.A.; Polombo, P.A.; Gruber, H.E.; Rimoïn, D.L.; Cohn, D.H.; Eyre, Am. J. Hum. Genet. 56, 388-395, 1995

A>Title: An RNA-splicing mutation (G-45IVS20) in the type II collagen gene (COL2A1) in a

A;Reference number: I38867; MUID:95150028; PMID:7847372

A;Accession: I38867

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 440,'G',442-456,'E',458-480,'P',482-509 <TIL1>

A;Cross-references: EMBL:U15195; NID:g557053; PIDN:AAB60370.1; PID:g557054

R;Ramirez, F.

submitted to the EMBL Data Library, December 1988

A;Reference number: S04892

A;Accession: S04892

A;Molecule type: mRNA

A;Residues: 501-676,'A',678-783,'A',785-831,'PA',834,'F',836-1214 <RAM>

A;Cross-references: EMBL:X13783; NID:g30037; PIDN:CAA323030.1; PID:g930050

R;Vikkula, M.; Peltonen, L.

FEBS Lett. 250, 171-174, 1989

A>Title: Structural analyses of the polymorphic area in type II collagen gene.

A;Reference number: S05000; MUID:89325561; PMID:2753125

A;Accession: S05000

A;Molecule type: DNA

A;Residues: 630-640,'A',642-785 <VIK2>

A;Cross-references: EMBL:X16158; NID:g29951; PIDN:CAA34278.1; PID:g1335018; PIDN:CAA3427

PIN:CAA34283.1; PID:g1335023; PIDN:CAA34284.1; PID:g1335024

R;Bogaert R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoïn, D.L.; Cohn, D.H.; Eyre, D. J. Biol. Chem. 267, 22522-22526, 1992

A>Title: An amino acid substitution (Gly853--Sglu) in the collagen alpha 1(II) chain pro

A;Reference number: A44309; MUID:93054548; PMID:1429602

A;Accession: A44309

A;Status: nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: DNA; mRNA

A;Residues: 752-831,'PA',834,'F',836-1005,'K',1007-1036,'Q',1038-1052,'E',1054-1068,'T',

A;Cross-references: GB:I00977; NID:g18012; PIDN:AAB23914.1; PID:g258774

A;Note: sequence extracted from NCBI backbone (NCBI:117273); parts of this sequence we

A;Note: this translation is not annotated and this publication is not cited in GenBank

A;Note: mutant sequence associated with perinatal lethal hypochondrogenesis

R;Tiller, G.E.; Rimoïn, D.L.; Murray, L.W.; Cohn, D.H.

Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990

A>Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individua

A;Reference number: S16502; MUID:90251862; PMID:2339128

A;Accession: S16502

A;Molecule type: DNA

A;Residues: 1164-1184,'GPSKGDGANGIPGI',1185-1199 <TIL2>

A;Cross-references: EMBL:M37126; NID:g180808; PIDN:AAA52037.1; PID:g180809

A;Note: mutant sequence from a patient with spondyloepiphyseal dysplasia

R;Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.

Proc. Natl. Acad. Sci. U.S.A. 88, 2555-2559, 1995

A>Title: Identification and characterization of the human type II collagen gene (COL2A1)

A;Reference number: A02858; MUID:85190534; PMID:3857598

A;Accession: A02858

A;Molecule type: DNA

A;Residues: 1032-1056,'N',1058-1068,'T',1070-1487 <CHE>

A;Cross-references: GB:J00116; NID:g180395; PIDN:AAA51997.1; PID:g180396

R;Elima, K.; Vuorio, T.; Vuorio, E.

Nucleic Acids Res. 15, 9499-9504, 1987

A>Title: Determination of the single polyadenylation site of the human pro-alpha-1(II)

A;Reference number: A27280; MUID:89067771; PMID:2825137

A;Accession: A27280

A;Molecule type: DNA; mRNA

A;Residues: 1175-1487 <ELI>

A;Cross-references: EMBL:X06268; NID:g30096; PIDN:CAA29604.1; PID:g30097

A;Experimental source: fetal epiphyseal cartilage

R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.

Biochem. J. 237, 923-925, 1986

A>Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.



Query Match	53.4%;	Score	2476;	DB 1;	Length	1487;			
Best Local Similarity	50.9%;	Pred. NO.	2.7e-116;						
Matches	503;	Conservative	45;	Mismatches	272;	Indels	168;	Gaps	17;

  

QY	1	GPPEGEGPTGLP	PPPPGERGQPSGRGPPGADGVAGPKGPA	GERGSGPGFAGPKSGPGEAGRP	60	
DB	225	GNGEGDGEVGS	GPMGPRGPPGPPGPKPGDGEAGKPK	KAGERGPPGQGARGGPTGTGLP	284	
QY	61	----	GEAGLPGAKGLTGSPP	-----GSDPP	-----DKTKTSPGPA	--- 90
DB	285	GVKHGKGYPEL	DAKGEAGAPGVKSGSGSPGPMGPRGL	PGERTGPAGAAGAR	344	
QY	91	GQDRRGPPGP	PGARGOAGVMGPP	-----GPKGAAGEPK	KAGERGVGPPGPAVGPA	141
DB	345	GNDGQPGAPGP	PPFVPGAGGPPGAPGAGKGEAGPTC	ARGPGECAQGPRGEPGTPGSPGA	404	
QY	142	----	GKGEAGCAQGP	PGPAGPAGERGEGPAGSGPGFGL	PGPAGPPGEACKPKGEQVP	195
DB	405	GASGNGPTDI	PKAKLSAGAPGATAGPPGPRGPPDPQA	TGTPGKQTKRPIAGFK	464	
QY	196	GD	-----LGAPGSPGAPGEGPTGLP	-----GPPGERGPGSGRGP	PGADGVA	237
DB	465	GEQGPKEGEP	AGPQAPGAPGAGEEGKRGARGEP	GGVGTGPPFERGAPGNRGFP	QQGGLA	524
QY	238	GPKGPAGERGSG	PGPAGPKSGPGEAGRGPBGAGL	PKAKGLTGS	PGSGPGDPGKTPGPPAGD	297

RESULT 6

A41182  
collagen alpha 1(II) chain precursor - mouse  
C.Species: Mus musculus (house mouse)  
C.Date: 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 13-Aug-1999  
C.Accession: A41182; A44885  
R: Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.  
J. Biol. Chem. 266, 16862-16869, 1991  
A.Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and  
A.Reference number: A41182; MUID:91358489; PMID:1885613  
A.Accession: A41182  
A.Status: preliminary; not compared with conceptual translation  
A.Molecule type: DNA  
A.Residues: 1-1419 <MET>  
A.Cross-references: GB:M65161  
R: Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.  
Development 111, 945-953, 1991  
A.Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilage.  
A.Reference number: A44885; MUID:91347939; PMID:1879363  
A.Accession: A44885  
A.Molecule type: DNA  
A.Residues: 1-28 <CHE>  
A.Cross-references: GB:S63190; NID:9234368; PIDN:AAB19627.1; PID:9234369  
A.Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBI:P:63192)  
C.Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology

C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trim  
F;1191-1419/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 53.0%; Score 2460; DB 2; Length 1419;  
Best Local Similarity 50.6%; Pred. No. 1.6e-115;  
Matches 500; Conservative 49; Mismatches 271; Indels 168; Gaps 17;  
Qy 1 GPGGPGCTGLPGRGPGSRGPGADGVAGPKAGRGSPGAPGKSGPCEAGRP 60  
Db 157 GNFGEPEGVGVGMPGPPGAPGPGDDGAGPKSGERGLPFGKAGFPETGLP 216  
Qy 61 -----GEAGLPKAGLTGSP-----GSPGP-----DGKTPGPPA--- 90  
Db 217 GVXGHRGYPGLDGAKEAGAPGVKSGSPGNGSPGMPGRLPCEGRGTGPACAAGAR 276  
Qy 91 GQGRGCPGPPGARGQAGVGMGFP-----GPKGAGEPKXAGRGVPPGCAVGA 141  
Db 277 GNDGQGPAGPPGVGAPGPGAPGAKGAGTGAKEGAGSGRGEFNPSPGPA 336  
Qy 142 GKDGE-----AGAQPPGAPGAGEGQGP-----AGSPGFQGLP 177  
Db 337 GASGNPDTGIPGAKGSAGAPGIAGAPFPFGPRGPPGQATGTLGPKQAGEPGIAGFK 396  
Qy 178 GPAGPCEACKPGEQGVGDLGAPGSPGAPGPGTGLPGRGGRGSGRFGCAQVA 237  
Db 397 GDGPGKGTGPAGPGAPGAPGAGGKARGEPGAGPIGPPGREGAGPFGQDGLA 456  
Qy 238 GPKGAGERGSPGAPGKSGPGEAGRPGEAGLPKAGLTGSPGSPGPGKTPGPPAGQD 297  
Db 457 GPKGAPGERGSPGLAGPKANGDPRGPGFGLPGARGLTGTRCDAGPKGVPSGAPGED 516  
Qy 298 GRGPGPPCARQAGVGMGFPFGPKGAGGPKGAGRGVPGCAVGPAGKDGEGAGAGPP 357  
Db 517 GRGPGPGQARQPGVGMGFPFGPKGANGEPGKAGSKLAGAPGLGLPKGQGTGAAGPP 576  
Qy 358 GPAGPAGERGEGQFAGSPGQGLPAGPFPBGAGPFGQGVGDLGAPGSPGAGE--- 413  
Db 577 GPGSPAGERGEGQAGPFGSGFQGLPAGPFPBGEGKQDQGLPGEAGAPGLVGRGERGFP 636  
Qy 414 -----PGTGLP-----GPGSRGPGSPGFGAGDVAGPKGA 447  
Db 637 GERGSPGAQGLGPRGLTGTDTGPKGAAGDPGPPGAGQPPGLQGMFGEAGAGIAGPK 696  
Qy 448 GERGSPGAPGKSGPGE-----AGRFGAGLPKAGLTGSPGSPGPD 489  
Db 697 GDRGDPGKPGAPKDGKGLTGTIPGPPGAGANGKGEVGPSPGSGTGARGAPCEP 756  
Qy 490 KTKTPGPAQDGERGPPGPPGAR-----GQAVMGFGPKGANGEP-----GKA 534  
Db 757 GETGPPGPAFAGPPGADGQPKAGQDQAGKQAGAPGPPGPGSGAPGPGPGTGTGTPK 816  
Qy 535 GERGVPPGPA-----VGPAGKDGAGAQGPPGPA-----GPAGERGEGQFAGSPGFQ 582  
Db 817 GARGAQPPGATGFPGAAGRVPPGANGNPAGFPAGPKDQKGVGDSGTPGAGDP 876  
Qy 583 GLPGAPGPAEAKRGEQGVGDLGAPGSPGAPGPGTGLPGRGGRGSGRFGPAD 642  
Db 877 GLQGPAGAPGKGEPPGDDGSGLDGPPGQGLAGQGVGLPQQRGERGFPGLPDPGEP 936  
Qy 643 GVAGPKGAPAGERGSPGAPGKSGPGEAGRPBAGLPFA-----KGLT 684  
Db 937 GKQAGPAGSGDRGPPGVPVPPGLTGPAGEPFGREGSPGADGPPRDGAAGVKGDRGTAL 996  
Qy 585 GSPGSGPDGKTPGPP-----AGODRPPGPPGAGQAGVGMGFPGPKCAAGEP 735  
Db 997 GARGAPGPPGSPGAPGTSKQDRGEAGAQGMGSGSPGAGARGIAGPPGPGDKGSGEQ 1056  
Qy 736 GK---AGERGVPPGPAVGPAGKDGAGAGQPPGAPGAGERGEGQFAGSPGFGGLPFA 792  
Db 1057 GERGLKGRFTGLQGLPFPFPSPGQAGSGAPGSGPRGPPGVPSPGKSGNGIPGI 1116  
Qy 793 GPPGEGAKPGEQGVGDLGAPGSPGAG 820

Db 1117 GPPGRGRSGETGVPVGPSPGSPGPPGP 1144

## RESULT 7

B41182  
collagen alpha 1(II) chain precursor (long splice form) - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 09-Jul-2004  
C;Accession: B41182  
R;Metaeairant, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.  
J. Biol. Chem. 266, 16862-16869, 1991  
A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and  
A;Reference number: A41182; MUID:91358489; PMID:1185613  
A;Accession: B41182  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-1487 <MET>  
A;Cross-references: UNIPROT:Q62031; UNIPROT:Q62032; UNIPROT:Q62033; GB:M65161  
C;Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;  
C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trim  
F;33-91/Domain: von Willebrand factor type C repeat homology <VWC>  
F;1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 53.0%; Score 2460; DB 2; Length 1487;  
Best Local Similarity 50.6%; Pred. No. 1.7e-115;  
Matches 500; Conservative 49; Mismatches 271; Indels 168; Gaps 17;

Qy 1 GPPGPGTGLPGRGPGSRGPGADGVAGPKAGRGSPGAPGKSGPCEAGRP 60  
Db 225 GNFGEPEGVGVGMPGPPGAPGPGDDGAGPKSGERGLPFGKAGFPETGLP 284  
Qy 61 -----GEAGLPKAGLTGSP-----GSPGP-----DGKTPGPPA--- 90  
Db 285 GVXGHRGYPGLDGAKEAGAPGVKSGSPGNGSPGMPGRLPCEGRGTGPACAAGAR 344  
Qy 91 GQGRGCPGPPGARGQAGVGMGFP-----GPKGAGEPKKAGRGVPPGCAVGA 141  
Db 345 GNDGQGPAGPPGVGAPGPGAPGAGKAGTGAKEGAGSGRGEFNPSPGPA 404  
Qy 142 GKDGE-----AGAQPPGAPGAGEGQGP-----AGSPGFQGLP 177  
Db 405 GASGNPDTGIPGAKGSAGAPGIAGAPFPFGPRGPPGQATGTLGPKQAGEPGIAGFK 464  
Qy 178 GPAGPAGERGEGQFAGSPGQGLPAGPFPBGAGPFGQGVGDLGAPGSPGAGE--- 237  
Db 465 GDRGDPGKPGAPKDGKGLTGTIPGPPGAGANGKGEVGPSPGSGTGARGAPCEP 524  
Qy 238 GPKGAGERGSPGAPGKSGPGEAGRPGEAGLPKAGLTGSPGSPGPGKTPGPPAGQD 297  
Db 525 GPKGAPGERGSPGLAGPKANGDPRGPGFGLPGARGLTGTRCDAGPKGVPSGAPGED 584  
Qy 298 GRGPGPPCARQAGVGMGFPFGPKGAGGPKGAGRGVPPGCAVGPAGKDGEGAGAGPP 357  
Db 585 GRGPGPPQARQPGVGMGFPFGPKGANGEPGKAGSKLAGAPGLGLPKGQGTGAAGPP 644  
Qy 358 GPAGPAGERGEGQFAGSPGQGLPAGPFPBGAGPFGQGVGDLGAPGSPGAGE--- 413  
Db 645 GDSGPAQGLGPRGLTGTDTGPKGAAGDPGPPGAGQPPGLQGMFGEAGAGIAGPK 704  
Qy 414 -----PGTGLP-----GPPSRGPGSPGFGAGDVAGPKGA 447  
Db 705 GERGSPGAQGLGPRGLTGTDTGPKGAAGDPGPPGAGQPPGLQGMFGEAGAGIAGPK 764  
Qy 448 GERGSPGAPGKSGPGE-----GPRGEGAGLPKAGLTGSPGSPGPD 489  
Db 765 GDRGDPGKPGAPKDGKGLTGTIPGPPGAGANGKGEVGPSPGSGTGARGAPCEP 824  
Qy 490 KTKTPGPAQDGERGPPGPPGAR-----GQAVMGFGPKGANGEP-----GKA 534  
Db 825 GETGPPGPAFAGPPGADGQPKAGQDQAGKQAGAPGPPGPGSGAPGPGTGTGTPK 884  
Qy 535 GERGVPPGPA-----VGPAGKDGAGAQGPPGPA-----GPAGERGEGQFAGSPGFQ 582



QY 52 GSPGEAGRCEAGLPGAKGLTGPSPGPDGKGTGPPGAGDGRPGPPGPGAGGAGVM 111  
 Db 332 GLPGERGRGSSGAAGAR---GNDGLPGP---AGPPGVGPAGARFGFPGARSGKEAGPT 385  
 QY 112 GPPGPKAAGEPKKAGERGVPGP-----PGAVGPAGKDGAEAGAGPPGPPAGPA 159  
 Db 386 GARGPEGAQGPGRGSGTGPSPGSGASGNPGTGDGIPGAKSGSGASGIAGAPGFPGRGPP 445  
 QY 160 GERGEQGP-----AGSPGFQGLPGPAGPPGAGPGEQGVGPDGLGAPGSPGAGEPPT 213  
 Db 446 GPGQATGPLPGKETGDPGIAKGHEGPKGIGSAGPQAGPAGAGEGKGARGEPGAA 505  
 QY 214 GLPFPGERGGSGRFGPGADGVAGPKGPPAGRSGFPGAGPKGSGEAGRGEAGLPGAK 273  
 Db 506 GPLGPPGERGAPGNRGFPQDGLAGPKGAPGREGVPLGGLGKGGNDGPRGEPGLPGAR 565  
 QY 274 GLTSGSGSPGDKTGPAGODGRGPPGPPGAGGAGVMPGPPGPKGAAGEGPKAGER 333  
 Db 566 GLTGPRGDAGPGKVGSPGAAGEDGRPPGPGQARGQFVWVGPPGPKGANGEFCKAGEK 625  
 QY 334 GVPGPAGVGPAGKDGAEAGQPPGPPAGPAGERGQGPAGSPGFQGLPGPAGPPGEAGKP 393  
 Db 626 GLGAPGLRGLPKDGETGAQGNPGPAGPAGERGEQGPFGSGFQGLGPPGSGEGGKP 685  
 QY 394 GEQGVPLDGLAPGSPGAGE---PQPTGLPDPGRRGPPGRRGPPGADGVAGPKGAPAGER 450  
 Db 686 GDQGVFGEAGAGLVPRGERGFPGERGSGPQGLQSGRGLPGTGTGDPKAGATGSPGPN 745  
 QY 451 GSPGPAGPKGSPCEAGRGEAGLPGAKGLTGPSPGPDGKGTGPPGAGDGRGPPGPP 510  
 Db 746 GAQGPGLQGNFGEAGAGISQPKDRDGTGKPEGAPGDKGSRGLTGLPPEPAGPN 805  
 QY 511 GARGQAGVWGPFPKGAAGEPKKAGERGVPPGPPGAVGPAKD-----GEA 555  
 Db 806 KEKSGSGSPGPGIVGARGAPCDRGENGPPGPPGAGPAGPPGSDGQAGLKGQSGQKXDA 865  
 QY 556 GAQGPAGPA-----GPAGERGEQGPAGSPGFQGL-----PGRAGPP 591  
 Db 866 GAGPQGSAGPAGPGPTGVNPGKARGAQGPAGATGPPGAAGRVGTGPNGNPPGPP 925  
 QY 592 BEAGKPGQGVPLDGLAPGSPGAGEPPTGLPFPGERGPPGRRGPPGADGVAGPKGPA 651  
 Db 926 SSAGKEGPKVGRD---AGPTGRAGDPLQGPAGAPGKPEGEDGSPGDPGSGPQGLS 982  
 QY 652 GERGSPGAPKGS-----PGEAGRGEAGLPGAKGLTGPSPGPPGDPGKGTGPPGAGQ 705  
 Db 983 GNRGIVLPGQRGERGFPGLPSPGEPKQGGSGSDRPPGPPGPPGLTGPAGETGRE 1042  
 QY 706 GRPDPGPPGARGQAGVWGFPPKGAAGEPKKAGERGVPPGPPGAVGPAKDGEAGAGQPP 765  
 Db 1043 GNPSDGPDPGDRGATGINKDRGETGPIGAPAPSGPAPFVGTGTGQDRGSGSGPL 1102  
 QY 766 GPAGP-----AGERGEQGPAGSPGFQGLP-----SGPAG 789  
 Db 1103 GPSGPAGAGLNGPQGPGRDKGEAGEAGERGKQHRGFTGLQGLPPTGATGQATGSA 1162  
 QY 790 -----GPAGPPGAGKPGEGVPLDGLAPGP-----SGPAG 820  
 Db 1163 GPSGPRGPPGVGSGKDSNGIIFGPIGPPGPRGRSETGPAG 1205  
 RESULT 10  
 CGB07S  
 collagen alpha 1(III) chain - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 04-Dec-1986 #sequence revision 04-Dec-1986 #text change 09-Jul-2004  
 C:Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946  
 R:Fietzek, P.P.; Allmann, H.; Kauterberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979  
 A:Title: The covalent structure of calf skin type III collagen. I. The amino acid sequence  
 A:Reference number: A02862; MUID:80026026; PMID:488906  
 A:Accession: A02862  
 A:Molecule type: protein

A:Residues: 1-242 <PIE>  
 A:Cross-references: UNIPROT:P04258  
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979  
 A:Title: The covalent structure of calf skin type III collagen. II. The amino acid sequence  
 A:Reference number: A38001; MUID:80026027; PMID:488907  
 A:Accession: A38001  
 A:Molecule type: protein  
 A:Residues: 243-422 <DEW1>  
 R:Bentz, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979  
 A:Title: The covalent structure of calf skin type III collagen. III. The amino acid sequence  
 A:Reference number: A38002; MUID:80026028; PMID:488908  
 A:Accession: A38002  
 A:Molecule type: protein  
 A:Residues: 423-571 <BEN>  
 R:Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979  
 A:Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequence  
 A:Reference number: A38003; MUID:80026029; PMID:488909  
 A:Accession: A38003  
 A:Molecule type: protein  
 A:Residues: 572-808 <LAN>  
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979  
 A:Title: The covalent structure of calf skin type III collagen. V. The amino acid sequence  
 A:Reference number: A38004; MUID:80026030; PMID:488910  
 A:Accession: A38004  
 A:Molecule type: protein  
 A:Residues: 809-947 <DEW2>  
 R:Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979  
 A:Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequence  
 A:Reference number: A38005; MUID:80026031; PMID:488911  
 A:Accession: A38005  
 A:Molecule type: protein  
 A:Residues: 948-1049 <ALL>  
 A:Experimental source: skin  
 R:Henkel, W.  
 Biochem. J. 318, 497-503, 1996  
 A:Title: Cross-link analysis of the C-telopeptide domain from type III collagen.  
 A:Reference number: S71946; MUID:96404897; PMID:8803038  
 A:Accession: S71946  
 A:Molecule type: protein  
 A:Residues: 87-106;1017-1029;1037-1049 <HEN>  
 C:Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are H  
 C:Comment: The type III collagen molecule is a trimer of identical chains, linked to each  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline  
 F:1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>  
 F:1-14/Region: amino-terminal nonhelical telopeptide  
 F:15-1040/Region: helical  
 F:587-589/Region: cell attachment (R-G-D) motif  
 F:752-754/Region: cell attachment (R-G-D) motif  
 F:875-877/Region: cell attachment (R-G-D) motif  
 F:878-880/Region: cell attachment (R-G-D) motif  
 F:935-937/Region: cell attachment (R-G-D) motif  
 F:1041-1049/Region: carboxyl-terminal nonhelical telopeptide  
 F:195,107,119,938,950/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:197,950/Modified site: allysine (Lys) #status predicted  
 F:107/Binding site: carbohydrate (Lys) (covalent) #status experimental  
 F:1040,1041/Disulfide bonds: interchain #status predicted

Query Match 51.3%; Score 2379; DB 1; Length 1049;  
 Best Local Similarity 49.8%; Fred. No. 1.4e-111;  
 Matches 479; Conservative 48; Mismatches 250; Indels 144; Gaps 18;  
 QY 1 GPPGPPGTGLPFPGERGPPGRRGPPGSGRGF-----PGADGVAGPKGAPAGER 42  
 Db 78 GRPGRGPRGPPGPPGPMKGPAGMPGPFPMKGRHGRGFDGRNGEKGPAGLKGNGVFGED 137  
 QY 43 GSPGAPKGSFGGAGRPGEAGLPGAKGLTGPSPGPDGKGTGPPGAGQDGRPPGPP 102

[illegible]

RESULT 11

RESOLUT I I  
 39856  
 collagen alpha 1(III) chain precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text\_change 09-Jul-2004  
 C:Accession: S59856; S62120; S16373  
 R:Toman, P.D.; de Crombrughe, B.  
 Gene 147, 161-168, 1994  
 A>Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA  
 A:Reference number: S59856; MUID:95011609; EMBID:7996795

Db 758 GP---RGPAGPIPPGP---ACQPCDKXGSGSPFLPIAGPGRGGGEGHGPAGP 811  
QY 565 -----GPAGERGQPGAGSPFQGLPAGPAGPGE---AGKPGEGVPGDLGAPGSPGA 615  
Db 812 GAFQNGEGFAGKERGAPGKGGGPPGAGTGGSGPAGPQGQVKGERSPPGPGTA 871  
QY 616 GEGPPTGLPQPPGREGGSGRFPFGADGVAGPKGAGREGSGPAGPKGSGEAGRPGE- 674  
Db 872 GFGGKGLFGPNNNGNPPFPFGAPGKDGPPGAGNCSGPNFGIAGPKGDAGQPGEK 931  
QY 675 -----ACLPKAKGLTSPGSPGPDGKTGP-----PAPAGQDGRP 708  
Db 932 GPPGAGPSPGSPGLGIAGLTGARGLAGPPGMPGPRGSPGPGQIKGESKPGASGNGER 991  
QY 709 GPPGP-----PGARGQAGVMPGPPGPKGAAGE---PGKAGRGVPPPP 747  
Db 992 GPPGQGLPQGTAGTAGEPRDGNPDSGQPGDRGSGPKGDRGNGSGFAGCAPGHPGPP 1051  
QY 748 GAVGPAGKDG-----AGAQGPPGPPGAPGAGERGEGPAGSGPFGGLPGPA 792  
Db 1052 GPVPSGKSGDRGETGPGAGPGAGPAGARGAPGQGRGDKGTGERTGSGNGIKHGRFP 1111  
QY 793 GPPGAGKPGEGVPGDLGAPGSPGAG 820  
Db 1112 GNPFPFGSGAAGHQAGTSGPAGPRG 1139  
RESULT 12  
CGHU7L  
collagen alpha 1(III) chain precursor - human  
N:Alternate names: procollagen alpha 1(III) chain  
C:Species: Homo sapiens (man)  
C:Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 09-Jul-2004  
A:Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90399  
R:Prockop, D.J.  
submitted to the EMBL Data Library, February 1989  
A:Reference number: S05272  
A:Accession: S05272  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1240, 'V', 1242-1466 <PRC>  
A:Cross-references: UNIPROT:P02461; EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058  
R:Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.  
Biochem. J. 260, 509-516, 1989  
A:Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human  
erences.  
A:Reference number: S04642; MUID:89350838; PMID:2764886  
A:Accession: S04642  
A:Molecule type: mRNA  
A:Residues: 1-1196 <ALA>  
A:Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058  
A>Note: the complete sequence is not shown  
R:Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.  
Gene 78, 255-265, 1989  
A:Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene  
A:Reference number: PE0011; MUID:89378752; PMID:2777083  
A:Accession: PE0011  
A:Molecule type: DNA  
A:Residues: 1-176 <BN>  
A:Cross-references: GB:M26939; NID:G180813; PIDN:AAA52040.1; PID:G180814  
R:Tomari, P.D.; Ricca, G.A.; de Crombrughe, B.  
Nucleic Acids Res. 16, 7201, 1988  
A:Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre  
A:Reference number: S01726; MUID:89303360; PMID:3405773  
A:Accession: S01726  
A:Molecule type: mRNA  
A:Residues: 1-170 <TM>  
A:Cross-references: EMBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:G30061  
A>Note: the authors translated the codon CAG for residue 154 as His  
R:Janeczko, R.A.; Ramirez, F.  
Nucleic Acids Res. 17, 6742, 1989  
A:Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.  
A:Reference number: S04887; MUID:89386015; PMID:2780304

A:Accession: S04887  
A:Molecule type: mRNA  
A:Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634,  
A:Cross-references: EMBL:X15332; NID:G29545; PIDN:CAA33387.1; PID:G930045  
A>Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide  
R:Seyer, J.M.; Kang, A.H.  
Biochemistry 16, 1158-1164, 1977  
A:Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide  
A:Reference number: A90399; MUID:77134724; PMID:557335  
A:Accession: A90399  
A:Molecule type: protein  
A:Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEV1>  
A:Experimental source: liver  
A>Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact  
R:Seyer, J.M.  
submitted to the Atlas, December 1977  
A:Reference number: A94562  
A:Accession: A94562  
A:Molecule type: protein  
A:Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>  
A:Experimental source: liver  
A>Note: author submitted corrections to A90399  
R:Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.  
Am. J. Hum. Genet. 53, 62-70, 1993  
A:Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual  
spring.  
A:Reference number: I51868; MUID:93304430; PMID:8317500  
A:Accession: I51868  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 186-194 <MIL>  
A:Cross-references: GB:S62925; NID:G386425; PIDN:AA013937.1; PID:G4261637  
R:Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.  
Biochem. J. 311, 939-943, 1995  
A:Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3  
A:Reference number: S59511; MUID:96067614; PMID:7487954  
A:Accession: S59511  
A:Molecule type: mRNA  
A:Residues: 302-423 <CHI>  
A:Cross-references: GB:S79877; NID:G1195576; PIDN:AA035615.1; PID:G1195577  
R:Seyer, J.M.; Kang, A.H.  
Biochemistry 17, 3404-3411, 1978  
A:Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr pe  
A:Reference number: A90414; MUID:79000343; PMID:687591  
A:Accession: A90414  
A:Molecule type: protein  
A:Residues: 399-675, 'N', 677-727 <SEY3>  
A:Experimental source: liver  
R:Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.  
J. Biol. Chem. 266, 5256-5259, 1991  
A:Title: G to T transversion at position +5 of a splice donor site causes skipping of th  
A:Reference number: I55349; MUID:91161621; PMID:1672129  
A:Accession: I55349  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 537-605 <LEE>  
A:Cross-references: GB:MS9312; NID:G180815; PIDN:AAA52041.1; PID:G180816  
R:Seyer, J.M.; Mainardi, C.; Kang, A.H.  
Biochemistry 19, 1583-1589, 1980  
A:Title: Covalent structure of collagen: amino acid sequence of alpha (III)-CB5 from ty  
A:Reference number: A90438; MUID:80198282; PMID:6246925  
A:Accession: A90438  
A:Molecule type: protein  
A:Residues: 728-895, 'A', 897-964 <SEY4>  
A:Experimental source: liver  
R:Lee, B.; Chiodo, A.A.; Lamanche, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Char  
J. Biol. Chem. 265, 17070-17077, 1990  
A:Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an  
A:Reference number: A38303; MUID:91009133; PMID:2145288  
A:Accession: A38303  
A:Molecule type: mRNA  
A:Residues: 861-1015 <COL>  
A:Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AA059383.1; PID:G









Db 237 GPPGPPGDPGMPGPTGSRGSPGPKPGCGEDGPGRNGNPGVEVFGAGSPGAGFFGAPGLP 296  
QY 34 GPK-----GPAGRGSP-----GPAGPKGSPGGRGPRGEGAGLGA 69  
Db 297 GLKGHRGKHGLEGKGEVGAQSGKEAGTGPWGAMGFLGPRGMPGGRGLGPGQAPQOR 356  
QY 70 GLTSGSPGSPGDKTGPDPAGQDGRP-----GPPGPPGARGAGVGMFFGPKGAA 120  
Db 357 GAHGMPPGPPGMPGLGIFGSSGFPNPNMKGEAGTGGARGEGGQGRGEGTGGPVGSP 416  
QY 121 GPPGKAGRGVFPGPAGVAGPKAGEAGAGAGQAPGPPAGPAGRGEGGQAGSPGFGPLGPA 180  
Db 417 GLPGAIGTGTGTPGPKGTGSPGTSPPGSGAGTGGSPGSGTGGQNSGLGDFGFKGEA 476  
QY 181 GPPGAGKPGEGQVPGDGLGAPGSPGAPGDPGTLGPPGGERGGGSGRFGAGVAGPK 240  
Db 477 GPKGPPGPHGIGQPIGPPGEEGKRGPRDPTGLGPPGVGERGAPGNRFFGSGDLGPK 536  
QY 241 GPAGRGSPGPPAGPKGSPGEGAGLPGAKGLTSGSPGPDGKTGPPGAPGQDGRP 300  
Db 537 GAQGERGPPVSSGPKGSGQDPRGPPGGLGARGLTGNPVGQEGKGLPLGAPGEDGR 596  
QY 301 GPPGPPGARGQAGVGMFFGPPGKAAGEPKAGRGVPPGPPGAVGAGKDGAGAGQPPGPA 360  
Db 597 GPPGSGIGIKGPGTGLFPGKSGNDPGKPGGAGNPGVPGQAGAPKDGKVGYPGPPGP 656  
QY 361 GPAGRGEGGPPGSPGFGQLGPPGAPGAGPKGQGVPGDLAGDPPSGP-----A 411  
Db 657 GLRGGERGQCPPTGTFQGHGPPPPGEGGKPGQGVGPPGAVGPPGPPGGERGNPGR 716  
QY 412 GPPGPTGLP-----GPPRGGPGGSRGPPGADGAVGAPKGPAGER 450  
Db 717 GEPGITGLPGEKMAGHGHPDGPKGSPGSPGTPGDTGPPGLQMGEGRGAGTGPCKDR 776  
QY 451 GS-----PCPAGPKGSPGAGRPGGAGLPGAKGLTSGPS---PGPD 489  
Db 777 GGIGKAGBTAGNDAGGLPLGLP---PGAGLLGKGBGPPRGLVGPFGSRGNPGR 833  
QY 490 GKTGP-----PPAGQDGRPPGPPGARGQAGVGMFFGPKGAAGEPKKAGERGVP--- 540  
Db 834 GENGPTGAVGFAGPQSGDQPGVKGEPPGPKGQDAGSPGQGLAGSPGHPGPNVPLK 893  
QY 541 -----GPPGAVGAPKAGEAGAP---PGPAGAGERGEGP-----AGSPGQGLP 585  
Db 894 GGRGTGPPGATGFPGSAGRVPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 953  
QY 586 GPAGP---PGEAGKPGEGVPGDLAGPSPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 642  
Db 954 GPAGPPGFGDKGDPGDCQPGDPGPPGAGTTGQGVGMFGQGRGMPGLGPACTP 1013  
QY 643 GVAGPKPAGRGSPGAPGKSGPGEAGPPGAGLPGAKGLTSGPSGPDGKTGTPGPA 702  
Db 1014 GKVGPTGATGDKGPPGPPGPNVPGPNVPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1073  
QY 703 ---GQDGRGPPGPPGARGQAGVGMFFGPKGAAGEPKKAGERGVFP----- 746  
Db 1074 GLPSSGAGTGPVAGAPDAGQGRDPPGGRGFIHLGRAGKGLGPGQPGRDKGDHGR 1133  
QY 747 -----PGAVGAPKDGAGAGQPPGPPGAPGAGERGEGQAGSPGFGSLGPPA 792  
Db 1134 GDRGQKGRHGTGLQGLPGLPPGPNBQSGAGTGPFPGRGPPGPPGPPGPPGPPGPPG 1193  
QY 793 GPPGEGAKPGE---QGVPGDLGAPGSPGAG 820  
Db 1194 GPPGVGRSGVGEAGPFGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1224

RESULT 14  
149607  
procollagen type V alpha 2 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 09-Jul-2004  
C:Accession: I49607

R:Andrikopoulos, K.; Suzuki, H.R.; Solursh, M.; Ramirez, F.  
Dev. Dyn. 195, 113-120, 1992  
A:Title: Localization of pro-alpha 2(V) collagen transcripts in the tissues of the deve  
A:Reference number: 149607; PMID:93214071; PMID:1297453  
A:Accession: I49607  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1497 <RES>  
A:Cross-references: UNIPROT:Q61431; GB:L02918; NID:G309180; PID:AAA37440.1; PID:G30918  
C:Genetics:  
A:Gene: Col5a-2  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology  
F:39-98/Domain: von Willebrand factor type C repeat homology <VWC>  
F:1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 49.6%; Score 2303.5; DB 2; Length 1497;  
Best Local Similarity 47.8%; Pred. No. 9.9e-108;  
Matches 473; Conservative 55; Mismatches 289; Indels 177; Gaps 19;  
QY 1 GPPGEP-----GPTGLGPPGE-----RGPSRSRFFGADGVA 33  
Db 235 GPPGEPGPPGMPGPIGSRGPPGPKPGEDGEGPRNGTGEVFGSGSPGARGFPFGAPGL 294  
QY 34 GPK-----CPAGERGSP-----GPAGPKGSPGGRGPRGEGAGLGA 69  
Db 295 GLKGHRGKHGLEGKGEVGAQSGKEAGTGPWGAMGFLGPRGMPGGRGLGPGQAPQOR 354  
QY 70 GLTSGSPGSPGDKTGPDPAGQDGRP-----GPPGPPGARGAGVGMFFGPKGAA 120  
Db 355 GAHGMPPGPPGMPGLGIFGSSGFPNPNMKGERGPHGARGEPGQGRGEGTGGPVGSP 414  
QY 121 GEPKAGRGVPPGPPGARGAGKAGAGQPPGPPGAPGAGERGEGQAGSPGFGPLGPA 180  
Db 415 GLPGAIGTGTGTPGPKGTGSPGTSPPGSGAGTGGSPGSGTGGQNSGLGDFGFKGEA 474  
QY 181 GPPGAGKPGEGQVPGDLAGPSPS---GPAGEGPTGLPGPGRGGRGPPGPPGADGVA 237  
Db 475 GERGPKGEPGPHIGPIGPIGPEEGKRGPRDGTGVPFGPMGERGAPGNRFFGSDGLP 534  
QY 238 GPKGAGERGSPGAPGKSGPGEAGLPGAKGLTSGSPSPGPDGKTGTPGPPAGQD 297  
Db 535 GPKGAQGERGPPVSSGPKGQGDPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 594  
QY 298 GRGPPGPPGARGQAGVGMFFGPKGAAGEPKKAGERGVPPGAVGAPKDGAGAGQGP 357  
Db 595 GRGPPGPPGARGQAGVGMFFGPKGAAGEPKKAGERGVPPGAVGAPKDGAGAGQGP 654  
QY 358 GPAGPAGGERGEGQAGSPGFGQLGPPGAPGPPGEGAGKPGEGQVPGDLAGPSPG 410  
Db 655 GPPGLAGERGEGAGPPGPTGFGQLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 714  
QY 411 --AGEGPTGLPDPGPPGEG---PGSRGPPGADGAVGAPKGA 447  
Db 715 GERGEPGTLGPKGKMGAGGHPDGPKNPGTGTGDTGPPGLQMGPPGGRGIAGTGP 774  
QY 448 GERGSPGAPGKSPGSE---AGRPGEAGLPGAKGLTSGSPSPGPDGKTGPP----- 495  
Db 775 GDRGIGEKAGBTAGNDAGRLGFLGPPGAPGLLAGPAGCEPGRGLVFPFSGRGNPGR 834  
QY 496 -----GPAGODGRPPGPPGARGQAGVGMFFGPKGAAGEPKKAGERGVP--- 540  
Db 835 GENGPTGAVGFAGPQSGDQPGVKGEPPGPKGQDAGSPGQGLAGSPGHPGHPGVLK 894  
QY 541 -----GPPGAVGAPKDGAGAGAGP---PGPAGPAGERGEGQAGSPGFGQLP----- 585  
Db 895 GGRGTQGPATGTPGSGAGRVPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 951  
QY 586 ---GPAGP---PGEAGKPGEGVPGDLAGPFGPPGAPGEPGTLGPPGPPGGRGPPGSGRFP 639  
Db 952 GDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1011  
QY 640 GADGAVGAPKGPAGERGSPGAPGKSGPGEAGRPGEGAGLPGAKGLTSGSPGPPGPKTGP 699

```

Db 1012 GTPKVGPTGATGDKGPPGVPVPGSGNPGVGPPEGPGAGNDGTGPRDCAVGGERGDRGP 1071
Qy 700 GPA---GQDRPGPPGPPGARGOAGVWGPFGKGAAGERGKAGRGVGPAGVAGPAGKD 756
Db 1072 GFAGLFGSQAGTTPGVPAGFDAGQORGEPPGSRGVPVGPPEGRAGKGLPGQPQPRDKGDN 1131
Qy 757 GE-----AGAGPPGPPAGPAGE-----RGEQPPAGSPGFGQLPGPA 792
Db 1132 GDRGDRGQKGRGFTGLQGLPPGPNNGEQSGAGIPGPPRGPVGGSSGKGNPGL 1191
Qy 793 GP-----PGEAGKPGQGVPGDLGAPGSPGAG 820
Db 1192 GPIGPPGVGSGVEAGPEGPPGPPGPPGPPG 1225

RESULT 15
A43291
collagen alpha 2(I) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A43291; A54328
R:Phillips, C.L.; Morgan, A.L.; Lever, L.W.; Wenstrup, R.J.
Genomics 13, 1345-1346, 1992
A:Title: Sequence analysis of a full-length cDNA for the murine pro alpha 2(I) collagen
A:Reference number: A43291; MUID:92372043; PMID:1505972
A:Accession: A43291
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1373 <PHI>
A:Cross-references: UNIPROT:001149; GB:X58251; NID:G50488; PIDN:CAA41205.1; PID:G50489
A>Note: sequence extracted from NCBI backbone (NCBI:112027)
R:Phillips, C.L.; Lever, L.W.; Pinnell, S.R.; Quarles, L.D.; Wenstrup, R.J.
J. Invest. Dermatol. 97, 980-984, 1991
A:Title: Construction of a full-length murine Proalpha2(I) collagen cDNA by the polymerase
A:Reference number: A54328; MUID:92084969; PMID:1748823
A:Accession: A54328
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-110 <PH2>
C:Genetics:
A:Gene: COL1A2
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: Coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:1145-1373/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 49.6%; Score 2300; DB 1; Length 1373;
Best Local Similarity 47.4%; Pred. No. 1.4e-107;
Matches 483; Conservative 50; Mismatches 275; Indels 210; Gaps 20;

Qy 1 GPPGPPGPTGLPPEGPGGPGSRFPFGADGVAGPKGAPAGRGSPGAPGKSGPCEAGRP 60
Db 103 GPRGPPGAVGAPGPGQFGPAGPEPQTGPAGRGPA---GSPGKAGEDGHGPKGRP 159
Qy 61 GAGLPAGKGLTSGSP-----GPDGKTGPPGAGQDGRPPGPPGARGQA 108
Db 160 GERGVVGPQAGKGFPTGLPFGKVGKSHGMDLKGQFQAQGVKGPAGENGTPCQA 219
Qy 109 GVMGFPKGAAGEPGKAGRGVPPGAVGAPGAKDGEA-----GAQGP-----P 153
Db 220 GARGLPGERGRVAGAPGAPAGSDGSGVGPAPGIPGAGPPGPPGKGLGPPVGNP 279
Qy 154 GPAGPAGERGE-----QGPAGSPGFG-----PCPA 180
Db 280 GPAGPAGPRGEVGLPGLSGVPPGPNPTNGLTGAAGATGLPVGAGAPGLGPPRGP 339
Qy 181 -----GPPGEAKGCEQVPGDLGAPGSPGAPGPTGLP----- 216
Db 340 GAAGATGARGLVGEPGPAAGSGKSGKNGKPEGSGVAGQPPGSGEGRKSGPGEAGSAGPA 399
Qy 217 GPPGERGPGSRGFTGADGVAGPKGAPAGRGSPGAPGKSGPAGRPGAGLPGAKGLT 276
Db 400 GPPGLRGSPGSRGLPGADGAGVWGPVGNRSGTGPAGIRGPNGDAGRPFGLMGRPLP 459

```

```

Qy 277 GSPGSPGPDGKTGPPGAPGODGRPPGPPGARGOAGVWGPFGKGAAGEPGKAGRGV 336
Db 460 GSFNVGSPSKRGFPVGLGIDGRPPGIFGAGPRKEAGNIGFPCKGKSGDPPKPERGHP 519
Qy 337 GPPGAVGAPAGKDEAGAAQAGPPGAPGAGRGEGGPPGAGSPGFGQLPGPAGPPGAGKPGSQ 396
Db 520 GLAGARGAPGPDGNNAGQAQPPGPGQVGGKGEQGPAGPPGFGQLPGPSTGTGEVGPGR 579
Qy 397 GVPDGLGAPGSPGAGE-----PGTGLPDP----- 422
Db 580 GLPGEFGLPGPAGPRGERGTPGESGAAGSPGTPGSRGSPGAPGPDGNNKGEAGAVAGP 639
Qy 423 -----PGERGG-----PGRGFPFGADGVAGPKPAGER 450
Db 640 GASGPGGLPGERCAAGIPGKGEKGETGLRGDTGNTGRDGARGIPGAVGAPGAPAGASGR 699
Qy 451 GS-----PGPAGPKSPGE-----AGRPGEAGLPKAK----- 477
Db 700 GEAGAGSPGAPGPRGSPGERGEVGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 759
Qy 478 GLTSGSPGPPDGKTGPPGAPGQDGRPPGPPGARGOAGVWGPFGKGA---GEPGKA 534
Db 760 GPTGSVGAAGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 819
Qy 535 GERGVGPPGAVGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 594
Db 820 GKEGIRGPRGDDQGVGTGTGASGPPGFV-----GEXGSPGEPGTAGAPTAGPQGL 873
Qy 595 KPEGEGGV-----GDLGAPGSPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 651
Db 874 GABGILGLPGRGERGLPGLIALGEPGLGSGPPGARGPPGAVGSP---GVNAPGEA 930
Qy 652 GERGSPGAPGKSGPGEAGRPGEAGLPKAGLITGSPGPPGPPGPPGPPGPPGPPGPP 711
Db 931 GRDNPSPGPPGPRDQGPQGHKGERGVPSIGPTGAAGAPGPHGSPGAPGKGNRGEPP 990
Qy 712 ---GPPGARGOAGVWGPFGKGAAGEPGKAGRGV-----GPPGAVGAPGAKDGEAG 762
Db 991 GSVGPPVAVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1050
Qy 763 GPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 820
Db 1051 GPVGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1108

```

Search completed: October 18, 2004, 13:42:59

Job time : 39.5431 secs

Aah63249 brachydan  
Q6uij5 brachydanio  
Aar24536 brachydan  
Q9w7c9 cynops pyrr  
Q6p422 xenopus tro  
Aah63191 xenopus t  
Q91717 xenopus lae  
Q7zt16 xenopus lae  
Q6pei9 brachydanio  
Aah58045 brachydan  
Q91718 xenopus lae  
Q7ztm4 xenopus lae  
Q6nz15 brachydanio  
Aah66384 brachydan

## ALIGNMENTS

## RESULT 1

CALL HUMAN STANDARD; PRT; 1464 AA.  
AC P02452; P78441; Q13896; Q13902; Q14037; Q14992; Q15176;  
AC Q15201; Q16050; Q7KZ30; Q7KZ34; Q81V15; Q9UML6; Q9UVM7;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Collagen alpha 1(I) chain precursor.  
GN Name=COL1A1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A., AND VARIANT SER-1434.  
RA Dalglish R.  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE OF 1-589 FROM N.A.  
RX PubMed=2843432;  
RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,  
RA Pretorius P.J.;  
RT "Complete nucleotide sequence of the region encompassing the first  
RT twenty-five exons of the human pro alpha 1(I) collagen gene  
RL (COL1A1).";  
RN [3]  
RP SEQUENCE OF 1-472 FROM N.A.  
RX MEDLINE=89025644; PubMed=3178743;  
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,  
RA Jaenisch R., Prockup D.J.;  
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain  
RT of human type I procollagen.";  
RL Biochem. J. 253:919-922(1988).  
[4]  
RP SEQUENCE OF 1-181 FROM N.A.  
RX MEDLINE=84270697; PubMed=6462220;  
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,  
RA Myers J., Williams C., Ramirez F.;  
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary  
RT conservation of a pattern of introns and exons.";  
RL Nature 310:337-340(1984).  
[5]  
RN SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.  
RP SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.  
RX TISSUE=Skin; PubMed=5529814;  
RN MEDLINE=71038625; PubMed=5529814;  
RA Click E.M., Bornstein P.;  
RT "Isolation and characterization of the cyanogen bromide peptides from  
RT the alpha 1 and alpha 2 chains of human skin collagen.";  
RL Biochemistry 9:4699-4706(1970).  
[6]  
RN SEQUENCE OF 425-1464 FROM N.A.  
RX MEDLINE=84080385; PubMed=6689127;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:23:10 ; Search time 146.729 Seconds  
(without alignments)  
3219.411 Million cell updates/sec

Title: US-10-658-989A-3  
Perfect score: 4640  
Sequence: 1 GPGGPGTGLPGPPGRRG.....GEOGVGDLGAPGSGPAGG 821

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_02.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2765	59.6	1464	1 CALL_HUMAN	P02452 homo sapien
2	2761	59.5	1461	2 Q76045	Q76045 homo sapien
3	2761	59.5	1464	2 Q8N473	Q8N473 homo sapien
4	2759	59.5	1460	1 CALL_CANFA	Q9xsj7 canis fami
5	2748	59.2	1453	2 Q63079	Q63079 rattus norv
6	2732	58.9	1453	1 CALL_MOUSE	P11087 mus musculu
7	2732	58.9	1453	2 Q810J9	Q810J9 mus musculu
8	2681	57.8	1069	2 Q6LAN8	Q6LAN8 homo sapien
9	2679	57.7	1069	2 CAA67261	CAa67261 homo sapi
10	2619	57.7	1453	1 CALL_CHICK	P02457 gallus gall
11	2568	55.3	1445	2 Q93251	Q93251 rana catesb
12	2561	55.2	1450	2 Q9V7B4	Q9V7B4 cynops pyrr
13	2523	54.5	1449	2 Q802B5	Q802B5 xenopus lae
14	2494	53.8	1225	2 Q6PCL3	Q6pcl3 mus musculu
15	2494	53.8	1225	2 AAh59281	Aah59281 mus muscu
16	2486	53.6	1160	2 Q14047	Q14046 homo sapien
17	2486	53.6	1487	2 Q14047	Q14047 homo sapien
18	2481	53.5	1487	2 Q77753	Q77753 canis fami
19	2479	53.4	1447	2 Q91B91	Q91b91 xenopus lae
20	2477	53.4	1418	2 Q28396	Q28396 equus cabal
21	2476	53.4	1418	1 CALL_HUMAN	P02452 homo sapien
22	2475	53.4	1420	2 Q90W37	Q90w37 gallus gall
23	2465	53.1	1269	2 Q7T227	Q7t227 gallus gall
24	2464	53.1	1442	2 Q62031	Q62031 mus musculu
25	2464	53.1	1442	2 Q62033	Q62033 mus musculu
26	2464	53.1	1459	1 CALL_MOUSE	P28481 mus musculu
27	2464	53.1	1459	2 Q62032	Q62032 mus musculu
28	2455	52.9	1419	2 Q80V13	Q80vy3 mus musculu
29	2454	52.9	1419	2 Q63123	Q63123 rattus norv
30	2449	52.8	1419	2 Q80X38	Q80x38 mus musculu
31	2430	52.4	1447	2 Q6P4U1	Q6p4u1 brachydanio

- RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,  
RA Prockop D.J.;  
RT "Nucleotide sequences of complementary deoxyribonucleic acids for the  
RT pro alpha 1 chain of human type I procollagen. Statistical evaluation  
RT of structures that are conserved during evolution.";  
RL Biochemistry 22:5213-5223(1983).  
RN [7]  
RN SEQUENCE OF 472-607 FROM N.A.  
RX PubMed=2981843;  
RA Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;  
RT "Multixon deletion in an osteogenesis imperfecta variant with  
RT increased type III collagen mRNA.";  
RL J. Biol. Chem. 260:691-694(1985).  
RN [8]  
RN SEQUENCE OF 488-625 FROM N.A.  
RX PubMed=3857621;  
RA Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;  
RT "Intron-mediated recombination may cause a deletion in an alpha 1 type  
RT I collagen chain in a lethal form of osteogenesis imperfecta.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874(1985).  
RN [9]  
RN SEQUENCE OF 1179-1464 FROM N.A., VARIANTS OI-II HIS-1277; ARG-1388 AND  
RP 1337-GLU-TYR-1338 DEL, AND VARIANT SER-1434.  
RX MEDLINE=93352646; PubMed=8349697;  
RA Chessler S.D., Wallis G.A., Byers P.H.;  
RT "Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)  
RT chain of type I collagen result in defective chain association and  
RT produce lethal osteogenesis imperfecta.";  
RL J. Biol. Chem. 268:18218-18225(1993).  
RN [10]  
RN SEQUENCE OF 1229-1454 FROM N.A.  
RC TISSUE=Bone; PubMed=3340531;  
RX MEDLINE=88424208; PubMed=8340531;  
RA Maekelae J.K., Raassina M., Virta A., Vuorio E.;  
RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide  
RT domain.";  
RL Nucleic Acids Res. 16:349-349(1988).  
RN [11]  
RN SEQUENCE OF 1-34 FROM N.A.  
RX MEDLINE=85130970; PubMed=2857713;  
RA Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;  
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.  
RT Promoter structure, AluI repeats, and polymorphic transcripts.";  
RL J. Biol. Chem. 260:2315-2320(1985).  
RN [12]  
RN SEQUENCE OF 1-34 FROM N.A.  
RX MEDLINE=8897389; PubMed=3480516;  
RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;  
RT "Regulatory elements in the first intron contribute to transcriptional  
RT control of the human alpha 1(I) collagen gene.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).  
RN [13]  
RN SEQUENCE OF 1-44 FROM N.A.  
RX MEDLINE=88033098; PubMed=2822714;  
RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,  
RA de Wet W.;  
RT "DNA sequences in the first intron of the human pro-alpha 1(I)  
RT collagen gene enhance transcription";  
RL J. Biol. Chem. 262:15151-15157(1987).  
RN [14]  
RN SEQUENCE OF 175-187 AND 274-289.  
RX PubMed=2169412;  
RA Baetge B., Notbohm H., Diebold J., Lehmann H., Bodo M., Deutzmann R.,  
RA Muller P.K.;  
RT "A critical crosslink region in human-bone-derived collagen type I.  
RT Specific cleavage site at residue Leu95.";  
RL Eur. J. Biochem. 192:153-159(1990).  
RN [15]  
RN SEQUENCE OF 263-268.  
RC TISSUE=Skin;  
RX MEDLINE=71001508; PubMed=4319110;  
RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;  
RT "A comparative study of glycopeptides derived from selected vertebrate  
collagens. A possible role of the carbohydrate in fibril formation.";  
J. Biol. Chem. 245:5042-5048(1970).  
RN [16]  
RN SEQUENCE OF 281-302; 402-420; 823-842; 924-944; 1026-1045 AND  
RP 1143-1162 FROM N.A.  
RX PubMed=2374517;  
RA Labhard M.E., Hollister D.W.;  
RT "Segmental amplification of the entire helical and telopeptide regions  
RT of the cDNA for human alpha 1 (I) collagen.";  
RL Matrix 10:124-130(1990).  
RN [17]  
RN SEQUENCE OF 425-490; 965-1024; 999-1039 AND 1453-1464 FROM N.A.  
RX MEDLINE=83064528; PubMed=6183642;  
RA Chu M.-L., Myers J.C., Bernard M.P., Ding J.-F., Ramirez F.;  
RT "Cloning and characterization of five overlapping cDNAs specific for  
RT the human pro alpha 1(I) collagen chain.";  
RL Nucleic Acids Res. 10:5925-5934(1982).  
RN [18]  
RN SEQUENCE OF 710-745 FROM N.A., AND VARIANT OI-II ARG-728.  
RX PubMed=2339700;  
RA Wallis G.A., Starman B.J., Zinn A.B., Byers P.H.;  
RT "Variable expression of osteogenesis imperfecta in a nuclear family is  
RT explained by somatic mosaicism for a lethal point mutation in the  
RT alpha 1(I) gene (COL1A1) of type I collagen in a parent.";  
RL Am. J. Hum. Genet. 46:1034-1040(1990).  
RN [19]  
RN SEQUENCE OF 746-781 FROM N.A., AND VARIANT OI-III SER-767.  
RX MEDLINE=95187161; PubMed=7881420;  
RA Forlino A., Zolzezi F., Valli M., Pignatti P.F., Cetta G.,  
RA Brunelli P.C., Mottes M.;  
RT "Severe (type III) osteogenesis imperfecta due to glycine  
RT substitutions in the central domain of the collagen triple helix.";  
RL Hum. Mol. Genet. 3:2201-2206(1994).  
RN [20]  
RN SEQUENCE OF 1187-1220 FROM N.A., AND VARIANT CYS-1195.  
RX PubMed=3170557;  
RA Cohn D.H., Apone S., Eyre D.R., Starman B.J., Andreassen P.,  
RA Charbonneau H., Nicholls A.C., Pope F.M., Byers P.H.;  
RT "Substitution of cysteine for glycine within the carboxyl-terminal  
RT telopeptide of the alpha 1 chain of type I collagen produces mild  
RT osteogenesis imperfecta.";  
RL J. Biol. Chem. 263:14605-14607(1988).  
RN [21]  
RN SEQUENCE OF 1440-1464 FROM N.A.  
RX MEDLINE=90110490; PubMed=2295701;  
RA Willing M.C., Cohn D.H., Byers P.H.;  
RT "Frameshift mutation near the 3' end of the COL1A1 gene of type I  
RT collagen predicts an elongated pro alpha 1(I) chain and results in  
RT osteogenesis imperfecta type I";  
RL J. Clin. Invest. 85:282-290(1990).  
RN [22]  
RN SEQUENCE OF 1454-1464 FROM N.A.  
RX PubMed=1995349;  
RA Maatta A., Bornstein P., Penttinen R.P.;  
RT "Highly conserved sequences in the 3'-untranslated region of the  
RT COL1A1 gene bind cell-specific nuclear proteins.";  
RL FEBS Lett. 279:9-13(1991).  
RN [23]  
RN REVIEW ON VARIANTS.  
RX MEDLINE=91184577; PubMed=2010058;  
RA Kuivaniemi H., Tromp G., Prockop D.J.;  
RT "Mutations in collagen genes: causes of rare and some common diseases  
RT in humans";  
RL FASEB J. 5:2052-2060(1991).  
RN [24]  
RN REVIEW ON VARIANTS.  
RX MEDLINE=97259599; PubMed=9101290;  
RA Kuivaniemi H., Tromp G., Prockop D.J.;  
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
RT associated collagen (type IX) and network-forming collagen (type X)  
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
RL Hum. Mutat. 9:300-315(1997).  
RN [25]

RP REVIEW ON OI VARIANTS.

RP REVIEW ON OI VARIANTS.  
RX MEDLINE=91374476; PubMed=1895312;

Query Match 59.6%; Score 2765; DB\_1; Length 1464;

Query Match	59.6%;	Score 2765;	DB.1;	Length 1464;
Best Local Similarity	55.4%;	Pred. No. 2.6e-103;		
Matches 561;	Conservative 31;	Mismatches 228;	Indels 192;	Gaps 19;

QY		1	GPBPFGPTGLPGRPGRGSGRFFPGADGVAGPKGAPAGERGSPGAPGKSGPGEA----	57
Db		179	GPMGSGSRGLGLFPGFPAGPGQGFQGGPFGEFGEFCASGMGPRGPPGFKNGDDGEAKP	238
QY		58	GRPGEAGLP---GAKGLTGSFGSCP-----DGKTGPGPAGDQRPGPPPPGAR	105
Db		239	GRPGERGPPQGARGLPGTTAGLDPMXKHGRFSGLDGAKGDAGPAGPKGEPGSPENGAP	298
QY		106	GQAGVMGFPGBKGAAGBFGKAGER-----GVPPFGAVGPA-----CKGGEAGAQ	150
Db		299	GQMGRGLPLGRGRFGAPGAPAGRGNDGATGAAGTFGPTGAPPPGFGVCAVGAKGEAGPQ	358
QY		151	GPPPGAPGAGEGEGQP-----AGSPGFQCLGPAA	180
Db		359	GRGSEGFQGVGRBFPPPGAGAAGPAGNFGAQCGFAKGANGAPCIAGAPFGPGARPS	418
QY		181	GPPGAGKFGQQVFQDLGAP-----GPSGA-----GEPGFT	211
Db		419	GFQPGCGPPKGNSEFGAPGSKDGTAKGEGPCPVGVQGPCPAGEEKRGARGECPGT	478
QY		214	GLPGBPGRGPGRSRFPFGADVAGPKGAPAGERGSGPGAPGKSGPGEAGRPOEAGLPKAK	273
Db		479	GLPGPPGRGPGRSRFPFGADVAGPKGAPAGERGSGPGAPGKSGPGEAGRPOEAGLPKAK	538
QY		274	GLTGSPGSGPDGKTGTPPGPADQDRGPPGPPGARQOAGVMGFPQPKCAACEPKKAGER	333
Db		539	GLTGSPGSGPDGKTGTPPGPADQDRGPPGPPGARQOAGVMGFPQPKCAACEPKKAGER	598
QY		334	GVPGPPGAVGPAKDGEAGAGQPPGPAPGAGERGEOGPAGSPGFQGLPGAPGPPGEAKP	393
Db		599	GVPGPPGAVGPAKDGEAGAGQPPGPAPGAGERGEOGPAGSPGFQGLPGAPGPPGEAKP	658
QY		394	GFQGVPGDLGAPGSGPAGE-----RGP-----TGLGPPP	423
Db		659	GFQGVPGDLGAPGSGPAGE-----RGP-----TGLGPPP	718
QY		424	GERGFGSRGFFGADVAGPKGAPAGERGSGPGAPKSGPGE-----AGRPEAGLP	474
Db		719	GSQAGPQLGWMPGERGAAGLPCKPKDGDAGPKADGSPGKGVRLGTPIGPPGPAGAP	778
QY		475	GAK-----GLTSGPSGPPGKTKTPGPAPAGDQRPGPPPPARQ-----	515
Db		779	GDKSGSGPGAPTGARCAPGDRGEFPFGPAGFAGPPGADQOPKAKEGPDGAKGDA	838
QY		516	-----AGVMGFPGPKGAAGEPKAGERVPQPCA-----VGPACKDGEAGAQPP	561
Db		839	GPFPAGPAGPPGPLNVAGAKAGSAGAPCATGFPCAAGRVPPFPSGNAGFPPPP	898
QY		562	GPAGP---AGERGEOGPAGSPFOGLPGPAPGPEGACKPGEQGVPGDLGAPGSPGAPGP	618
Db		899	GPAGKEGKGRGETGPAGRPGEVPPGPPGPAGEKSGCADGPAGAPTPPGQTIAQOR	958
QY		619	GTGLPPLPPGRRGSGSFFGADVAGPKGAPAGERGSGPGACP---KCSPEAGEPRGEA	675
Db		959	GVVLPPQKRGFRPLPGPGEFGKQOPSAGSERPPGFMGPPGLAGPPESGREGAP	1018
QY		676	GLPGAAGLTGSPSGPPDKTKTFFGAPAQDGRGPPGPPGARQAQWGMFPFKGAAGEP	735
Db		1019	GAEGSPGRDGSFPAKCDRGETGPAGPPGAPGAPGAPGVGPAKSGCDRGETGPAGPAPV	1078
QY		736	GTAGERGVPDPGAVGPAKKDGE-----AGAQQPQP-----AGPA	771
Db		1079	FPAAGARGPAGPCGPRGDKGETGEOGDRGIKHRGFGSLQGPFGPPGSGPGEQPSGASGPA	1136
QY		772	GERGBOGPAGSP---GFQGLPCPAGPCEAGKPCEQCVPGDLGAPGSPGAP	820

DB 1139 GFRPPGSGAPKGDGLNGLPGPIGGPGRGTGADGVGPDPGPPGPPGPPG 11

RESULT 2  
O76045 PRELIMINARY; PRT; 1461 AA.  
AC O76045  
ID O76045  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Pro alpha 1(I) collagen.  
DE Name=COL1A1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RX SEQUENCE FROM N.A.  
RP MEDLINE=85130970; PubMed=2857713;  
RA Chu M.L., de Wet W., Bernard M., Ramirez F.;  
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene:  
RT Promoter structure, AluI repeats, and polymorphic transcripts."; J. Biol. Chem. 260:2315-2320(1985).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=88329734; PubMed=2843432;  
RA D'Alessio M., Barnard M., Pretorius P.J., de Wet W., Ramirez F.;  
RT "Complete nucleotide sequence of the region encompassing the first  
RT twenty-five exons of the human pro alpha 1(I) collagen gene."; Gene 67:105-115(1988).  
RL [3]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=89025644; PubMed=3178743;  
RA Tromp G., Kulvaniani H., Stacey A., Shikata H., Baldwin C.T.,  
RA Jaenisch R., Prockop D.J.;  
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) c  
RT of human type I procollagen."; Biochem. J. 253:919-922(1988).  
RL [4]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=91138770; PubMed=1995349;  
RA Maatta A., Bornstein P., Penttinen R.P.;  
RT "Highly conserved sequences in the 3'-untranslated region of the  
RT COL1A1 gene bind cell-specific nuclear proteins."; FEBS Lett. 279:9-13(1991).  
RL [5]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=92157916; PubMed=1787829;  
RA Westerausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.  
RA Olsen A.S., Prockop D.J.;  
RT "Completion of the last half of the structure of the human gene fo  
RT the pro alpha 1 (I) chain of type I procollagen (COL1A1)."; Matrix 11:375-379(1991).  
RL [6]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=98107942; PubMed=9443882;  
RA Korikko J., Ala-Koriko L., De Paape A., Nytyinck L., Earley J.,  
RA Prockop D.J.;  
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and  
RT scanning by conformation-sensitive gel electrophoresis identifies  
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type  
RT identification of common sequences of null-allele mutations."; Am. J. Hum. Genet. 62:98-110(1998).  
RL [7]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=98107942; PubMed=9443882;  
RA Korikko J.M., Earley J.J., Nytyinck L., Depaepe A., Prockop D.J.,  
RA Ala-Koriko L.;  
RT Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.  
RL ENBL; AF017178; AAB94054.2;  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005737; C:cyclopasem; IEA.  
DR GO; GO:0005201; P:extracellular matrix structural constituent; IEA.  
DR GO; GO:0006817; P:phosphate transport; IEA.

DR InterPro: IPR008161; Clg helix.  
 DR InterPro: IPR008160; Collagen.  
 DR InterPro: IPR008885; Fib\_collagen\_C.  
 DR InterPro: IPR001007; VWF\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 18.  
 DR ProDom: PD000007; Clg helix; 3.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 DR SMART: SM00214; VWC; 1.  
 DR PROSITE: PS01208; WFC1; 1.  
 DR PROSITE: PS0184; WFC2; 1.  
 KW Collagen.  
 SQ SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30EA78E21 CRC64;

Query Match 59.5%; Score 2761; DB 2; Length 1461;  
 Best Local Similarity 55.3%; Pred. No. 3.8e-103;  
 Matches 560; Conservative 31; Mismatches 229; Indels 192; Gaps 19;

QY 1 GPPGPGTGLPGRGPGSRGPPGADGVAGPKGAGRGSPGAGPKGSGPGEA--- 57  
 DB 176 GPMGSGRGLRPGPGARPGQFGPGEPGPGASGPMGPRGPGPGKNGDDGEAGKP 235  
 QY 58 GRPGAGLP---GAKLTGSPSPGP-----DGTGPPGAGQGRGPPGPPGAR 105  
 DB 236 GRPGGPGPGQARGLFTAGLPMKGRGFGSLDGAKDAGPAGPKGEGSPGENGAP 295  
 QY 106 GQAGVMGPPGPKGAAGEPKAGER-----GVPGPPGAVGPA-----GKDGEGAGQ 150  
 DB 296 GQMGPRGLGEGRGPGAGPAGARGNDGATGAGPPGTGTPAGPPGFCVAGKAGEAGPQ 355  
 QY 151 GPPGAGPAGERGPGP-----AGSPGFGQLPQPA 180  
 DB 356 GPRGSEGGQVGRGEPGPGPAGAGPAGNPGADGQPGKANGAGPAGIAGAGFPGARGPS 415  
 QY 181 GPPGAGKPGEGVPGDLGAP-----GPSGPA-----GEPGPT 213  
 DB 416 GQGGPGPPPKNGSGEPGAPGSKDGTAKGEPGVGVGPPGPPGAGEGKRGAGEGPT 475  
 QY 214 GLPFGPGGPGSRGPFPGADGVAGPKGAGRGSPGAGPKGSPGEGAGRGEPGAGLPGAK 273  
 DB 476 GLPFGPGGPGSRGPFPGADGVAGPKGAGRGSPGAGPKGSPGEGAGRGEPGAGLPGAK 535  
 QY 274 GLTSGSPGPPGKTPGPCAGQDGRPPPPGARGQAGVMGPPGPKGAGEGPKAGER 333  
 DB 536 GLTSGSPGPPGKTPGPCAGQDGRPPPPGARGQAGVMGPPGPKGAGEGPKAGER 595  
 QY 334 GVPFGPGAVGPKDGEAGAGQPPGAGPAGERGQGPAGSPGQGLPGPAGPPGEGAGKP 393  
 DB 596 GVPFGPGAVGPKDGEAGAGQPPGAGPAGERGQGPAGSPGQGLPGPAGPPGEGAGKP 655  
 QY 394 GQGVPGDLGAPPGSGPAGE-----PGP-----TGLPGPP 423  
 DB 656 GQGVPGDLGAPPGSGAGERGFPGERGVGPPGPGPRGANGAPGNDKAGDAGAGAP 715  
 QY 424 GERGGPGSRGPGADGVAGPKGAGRGSPGAGPKGSPGE-----AGRGEGAGLP 474  
 DB 716 GSQAFGLQMPGEGKAGLPGPKDGRDAGPKADGSGGKDGVLGTGTPGPPGAGAP 775  
 QY 475 GAK-----GLTSGSPGPGDKTTPPGAGQDGRPPGPPGARGQ----- 515  
 DB 776 GSKGSGSPGAPGAPGAGCDRGEPPGPPGAGPAGPGADGQFGAGKEGPDGAGAGDA 835  
 QY 516 -----AGVMGPPGPKGAAGEPKAGERGVGPPGA-----VGPAGKDGEGAGAGPP 561  
 DB 836 GPPGAGPAGPPEP LGVAPKAGKAGSAGPPGATGPPGAGRVGPPGSGNAGPPGPP 895  
 QY 562 GPAGP---AGERGQGPAGSPGQGLPGPAGPPGEGAGPKGQGVGDLGAPGSPGAPAGE 618  
 DB 896 GPAGKEGKGRGPGTGTGAGRGVPPGPPGAGEKSGFADGAGAPGTPGPGQIAGOR 955  
 QY 619 GPTGLPGPPGPGSGRGGFGGADGVAGPKGAGRGSPGAPG---KSPGEGAGRGGEA 675

Db 956 GVVGLPQQRGERGFPGLPGSPGPKQKQPSGASGERGPPGMPGPPGLAGPPGSGREGAP 1015  
 QY 676 GLFGAKGLTGSPGSPGPDGKTGPPGAGQDGRPPGPPGAGQAGVMGFPDPKGAAGEP 735  
 Db 1016 GAEGSPGRDGSFGAKGDRGETGPPGPPGAPGAPGPPVGPAGKSGDRGETGPPGAPGV 1075  
 QY 736 GKAGERGVPPGPPGAVGAPGAGKGS-----AGAQPPGP-----AGPA 771  
 Db 1076 GPVGARGPAGPQGRDGTGEGQGRGKXGHRGFSGLGQFPGPPSPGQSPGSGASGPA 1135  
 QY 772 GERGQGPAGSP---GFGPLGPPGPPGAGKPKGEGQVPGDGLGAPSPGAP 820  
 Db 1136 GPRGPPGSAAGKGLGPGPIGPPGPRGTGDAGVPGPPGPPGPPG 1187

RESULT 3  
 Q8N473 PRELIMINARY; PRT; 1464 AA.  
 AC Q8N473;  
 DT 01-OCT-2002 (T-EMBLrel. 22, Created)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)  
 DE Alpha 1 type I collagen, preproprotein.  
 GN Names:COL1A1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Donald M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guraratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Murthy D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Krzywinski M.I., Skalska U., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 DR ENBL: BC035331; AAH3533.1; -;  
 DR GO: GO:0005581; C:collagen; IEA.  
 DR GO: GO:0005737; C:cyclopia; IEA.  
 DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR GO: GO:0006817; P:phosphate transport; IEA.  
 DR InterPro: IPR008161; Clg helix.  
 DR InterPro: IPR008160; Collagen.  
 DR InterPro: IPR008885; Fib\_collagen\_C.  
 DR InterPro: IPR009041; PMP\_SGCI.  
 DR InterPro: IPR001007; VWF\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 18.  
 DR ProDom: PD000007; Clg helix; 3.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 DR SMART: SM00214; VWC; 1.



FT MOD\_RES 1160 1160 3-hydroxyproline (By similarity).  
 FT CARBOHYD 261 261 O-linked (Gal...) (By similarity).  
 FT CARBOHYD 1361 1361 N-linked (GlcNAc...) (By similarity).  
 FT VARIANT 208 208 G -> A (in OI; severe).  
 SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;

Query Match 59.5%; Score 2759; DB 1; Length 1460;  
 Best Local Similarity 55.9%; Pred. No. 4.5e-103;  
 Matches 559; Conservative 31; Mismatches 230; Indels 180; Gaps 19;

QY 1 GPPGPGPTGLPCCPCEGPGS-----RGPFGA-----DGVAGPKPAGERGSGPPA 48  
 Db 187 GPPGAPGPGFQPGCEPCEPCASGPMGRPPGPPGKNGDDGEAGKGRGERGPPGQ 246  
 QY 49 GPKGPGEGARP-----GEAGLPGAKGLTGSFSGSPGPKTGPDPGAGQD 93  
 Db 247 GARGLFTAGLPGMKHGRFSLDGAAGDAGPAGKBPSPGNGAPGQMGPRGLPGER 306  
 QY 94 GRGPGPPGARGQ---AGWMGPPGPKGAAGB---PGKAGERGVDPGPAVGPAGKAGEA 147  
 Db 307 GRPGAPGACAKNGNDGATGAAGPPPTGAPGPPFGAVGAKGERAGPQAGSGSPGQVR 366  
 QY 148 GAQGGPPGAPGAGERGEQGP-----AGSPGFQGLPGRAGPGEAGKPGEQ 192  
 Db 367 GEPGPPGAGAGPAGNPGADQPGAKGANGAPGAGPAGPFGPAGPSPGQSPGPGPK 426  
 QY 193 GVPGLDGLAP-----GPSGA-----GPGPTGLPGERGGP 225  
 Db 427 GNSGPGAPNGKDTGAKBPPTGTPGPPGAGEGRGARGPPTGLPFPGERGGP 486  
 QY 226 GSRGPPGADVAGPKPAGERGSPGAPGKSPGEGAGPFGAGLPGAKGLTGSFGSPGD 285  
 Db 487 GSRGPPGADVAGPKPAGERGSPGAPGKSPGEGAGPFGAGLPGAKGLTGSFGSPGD 546  
 QY 286 GKTGPPGAGQDRGPPGPPGARGQAGVMGPPGKGAAGPFGKAGRGVDPGPAVGA 345  
 Db 547 GKTGPPGAGQDRGPPGPPGARGQAGVMGPPGKGAAGPFGKAGRGVDPGPAVGA 606  
 QY 346 GKDAGAGAGPPGAPGAGERGEQGPAGSPGQGLPAGPAGPKAGEQGVGDLGAP 405  
 Db 607 GKDAGAGAGPPGAPGAGERGEQGPAGSPGQGLPAGPAGPKAGEQGVGDLGAP 666  
 QY 406 GFSGPAGE-----PGP-----TGLPDPGGERGGSGRFP 435  
 Db 667 GFSGARGGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 726  
 QY 436 GADGAGPKPAGERGSPGAPGKSPGEGAGLPGAKGLTGSFGSGPGE-----AGRPGEAGLPGAK----- 477  
 Db 727 GERGAAGLPGKGRDGDGPKGADGSPKDGVRGLTGTGIPGPPGAPGPDGEGAGSGPA 786  
 QY 478 GLTSGFSGPDGKTPGPPGAGQGRPPGPPGARGQ-----AGVMGFP 522  
 Db 787 GPTGARGAPDRGEPGPPGAPGAGPAGADGQPKAGBPBGADAGKGDAGPAGPAGTGP 846  
 QY 523 GPKGAAGPCKAGERGVCPGCA-----VGPAGKDGAGAGQCPGPPAGP---AGER 570  
 Db 847 GPTGNVAGPPKARGSGAPGATGFPAGAGRVGPPGSGNAGPPGPPGAPGKGGKAR 906  
 QY 571 GQGGPAGSPGFLPFPAGPGEAGKPGEQGVGPDLAGPSPGAPGPPGTGLPFPGER 630  
 Db 907 GETGAPRGVEGPPGPPGAGEKSPGADGAPAGPPTGQIAGGVVGLPQGRGER 966  
 QY 631 GPGSGRFPAGDVAGPKPAGERGSPGAPG---KSPGAGRPAGLPGAKGLTGP 687  
 Db 967 GPPGPPGSGEGKPGSGTSGERPPGPMGPPGLAGPPGSGEGSGPAGSGSPGSDGSP 1026  
 QY 688 GSPGPDGKTPGPPGAGQDRGPPGPPGARGQAGVMGPPGKGAAGPFGKAGERGVGPP 747  
 Db 1027 GPKDRGETGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 1086  
 QY 748 GAVGAPGKDE-----AGAQPDPG-----AGPAGERGEGPAGSP 783  
 Db 1087 GPRGDKGTGQDGRGKGRHRSGLQFPGPPGSPGQSPGSGASGAPGPPGSGAGSP 1146

QY 784 ---GFGGLPGLPAGPPGAGKPGEQGVGPDGLGAPGSGGAPG 820  
 Db 1147 GKDLNGLPGLPAGPPGPRGRTGDAGVGPFPFPFPFP 1186

## RESULT 5

Q63079 PRELIMINARY; PRT; 1453 AA.  
 AC Q63079;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)  
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)  
 DE Collagen alpha1 (Fragment).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Bone, and Tooth;  
 RX MEDLINE=9916394; PubMed=10065941;  
 RA Brandsten C., Lundmark C., Christersson C., Hammarstrom L., Wurtz T.;  
 RT "Expression of collagen alpha1(I) mRNA variants during tooth and bone  
 formation in the rat";  
 RL J. Dent. Res. 78:11-19(1999).  
 DR EMBL: Z78279; CAB01633.1; -;  
 DR GO: GO:0005581; C:collagen; IEA.  
 DR GO: GO:0005737; C:cytoplasm; IEA.  
 DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR GO: GO:000617; P:phosphate transport; IEA.  
 DR InterPro: IPR008161; Clg helix.  
 DR InterPro: IPR008150; Collagen.  
 DR InterPro: IPR000885; Fib collagen\_C.  
 DR InterPro: IPR009041; PMP\_SGCI.  
 DR InterPro: IPR001007; VWF\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 18.  
 DR ProDom: PD000007; Clg helix; 3.  
 DR ProDom: PD002078; Fib collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 DR SMART: SM00214; VWC; 1.  
 DR PROSITE: PS01208; VWF\_C\_1; 1.  
 DR PROSITE: PS0184; VWF\_C\_2; 1.  
 KW Collagen.  
 FT NON TER.  
 SQ SEQUENCE 1453 AA; 137886 MW; E6896BDC19A4A1D8 CRC64;

Query Match 59.2%; Score 2748.5; DB 2; Length 1453;  
 Best Local Similarity 52.3%; Pred. No. 1.2e-102;

Matches 556; Conservative 33; Mismatches 231; Indels 243; Gaps 19;

QY 1 GPPGE---PGPTGLPGERGSGRPPG-----ADGVAGPKPAGER 42  
 Db 117 GPPGQDGIQGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLP 176  
 QY 43 GSPGAPGPKSGPGEAGRPGEAGLPGAKGLTGSFSGSPGPDGKTPGPPGAGQDRP---GPP 99  
 Db 177 GLPFPGAPGPPGQFPQGPGEPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 236  
 QY 100 GPPGARGQAGVMGFP-----GPKGAAGPCK-----A 126  
 Db 237 GQAGARGLPTAGLPGMKHRCFSLDGAKGDTGAPGKGPSPGSENGTQGMQPRGLP 296  
 QY 127 GSRGVGPPGAVGPAKAGEAGAQPPGP-----AGPAGERGEQGP- 167  
 Db 297 GERGRGPPGTAGAGNDGAVGAAGPPTGPTGTPGPPFGAAGKAGEAGPQARGSEGP 356  
 QY 168 -----AGSPGFQGLPAGPAGPAGCKP 189  
 Db 357 GVRGEPGPPGACAGPAGNPGADQPKAGKANGAPGAGPAGPAGPAGPAGPAGPAGPAG 416  
 QY 190 GEQGVGPDGLAP-----GPSGA-----GEPGPTGLPGER 222







```
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008885; Fib_collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; C1g_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS0184; VWF_C_2; 1.
DR Collagen.
KW SEQUENCE
SQ SEQUENCE 1453 AA; 138032 MW; 0B7F06BBB9A1D5EA CRC64;

Query Match 58.9%; Score 2732.5; DB 2; Length 1453;
Best Local Similarity 54.7%; Pred. No. 5,2e-102;
Matches 547; Conservative 33; Mismatches 240; Indels 180; Gaps 16;

QY 1 GPPGPGTGLPDPGE-----RGPGSRGPPGADGVAGPKGAPAGERGSGPPA 48
DB 180 GPPGAPGQGGPQGGPGEPPGPMGPPGPPGKNGKGGDAGKPRGRGPPGPPQ 239
QY 49 GPKSGPGEAGRP-----CEAGLPKAGLTSPPSGPPGDKTGPFGPAGQD 93
DB 240 GARGLPGTAGLPGMKHGRFGLDGAADGAPKGPSPFGNGAPGQMGRLPGER 299
QY 94 GRPGPPGPPGARGQAGVGMFPFKGAAGE-----PGKAGERGVGPPFGAVGPAKDGEA 147
DB 300 GRPGPPGTAGRGNDGAVGAAGPPGPTGTPGPPFGAVGAKGAGPQARGSGPQVR 359
QY 148 GAQGGPPGAPAGERGEGP-----AGSPGQGLPAGPPGEGEAGKGEQ 192
DB 360 GEPGPPGAGAGPAGNPGADQPGKAGANGAPGTAGAPFGARGSPGPGSGPPGPK 419
QY 193 GVPGLDGLAP-----GPSGPA-----GRPGTGLPDPGPPGPP 225
DB 420 GNSGEPGAPGNKDTGAKGEGCATGVQPPFAGEEKGARGPFGSLGPPGPPG 479
QY 226 GSRGPPGADGVAGPKGAPAGERGSPGAPKSPGAPGPPGAGLPKAGLTSPPSGPP 285
DB 480 GSRGPPGADGVAGPKGSPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 539
QY 286 KHTGPPGAGDGRGPPGPPGARGOAGVMGPPGKGAAGBPAGKAGRGVPPGAVGA 345
DB 540 KHTGPPGAGDGRGPPGPPGARGOAGVMGPPGKGAAGBPAGKAGRGVPPGAVGA 599
QY 346 GKDGEAGQGGPPGAPAGERGEGQGPAGSPGQGLPGPAGPPGAGKFGEGQVPCDLGAP 405
DB 600 GKDGEAGQAGPAPAGERGEGQGPAGSPGQGLPGPAGPPGAGKFGEGQVPCDLGAP 659
QY 406 GPSGAGE-----PP-----TGLPGERGPPGSRGFP 435
DB 660 GPSGARGEPFPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 719
QY 436 GADGVAGPKGAPAGERGSPGAPGKSPGE-----AGRPGEAGLPKAGLTSPPSG 486
DB 720 GEGGAGLPKPKGDRGDAGPKADGSPKDGARGLTGIPGPPGAPAGGDKGAPSGPP 779
QY 487 GPD-----GKTGPPGAGQDGRPPGPPGARGQ-----AGVMGFP 522
DB 780 GPTGARGAPDRGEAGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 839
QY 523 GPKGAAGEPKAGRGVPPGCA-----VGPAGKDGAGAGGPPGAPG-----AGER 570
DB 840 GPGVAGPAPKPPGAGPPGATGTPGAGRVGPPGSGNAGPPGPPGPPVKGEGKGR 899
QY 571 GEQGPAGSPGGLFPGAPGPPGAGKPGEQVQGVGDLGAPGSPGAPGPPGPPGPPG 630
DB 900 GETGAPRGVEGPPGPPGAPGKSPGADGAPGAPGPPGPPGPPGPPGPPGPPGPP 959
```





Db 534 GSPGPDGKTGPPGAGDGRPGAGPGGARGQAGVMGPPGPKGAAGEPKGERGAGP 593

Qy 340 GAVGAGKDGAGAGAGPPGAGPAGERGEOGAGSGFQGLPGAGPPGEGAKGEGQV 399

Db 594 GAVGAGKDGAGAGAGPPGAGPAGERGEOGAGSGFQGLPGAGPPGEGAKGEGQV 653

Qy 400 GDLGAPGSGPAGE-----PGP-----TGLPGPPGEGGP 429

Db 654 GNAGAPGAGARGEGFGERGVQPPQPPGPRGANGAPGNDGAGKAGDAGAGP 713

Qy 430 GSRGPPGADGVAGPKGAPGERGSGP-----AGKSGPGEAGRP---GEA 471

Db 714 GLEGMFGERGAAGLPGKAGDGPFGKAGDGPAGKDLGLTGTGTPGAGAGP 773

Qy 472 GLPGAKGLTSGSPGDPGKTPGCPA-----GQDGRGPP 507

Db 774 GPPGAGTGAAGFDRGEPGPPGAGPAGPPADGQPGAKGTGDAKAGDAGP 833

Qy 508 GPPGARGQAGVMGPPGPKGAAGEPKAGERGVPVPGAVGPAKDGEAGAGP 567

Db 834 GTCAGPAGZVGAAPGKAGSAGPPGATGPGAAAGRVGPPGSGNIGLPGPPG 893

Qy 568 GB---RQSGPAGSGFQGLPGAGPCEAGKPGEGQVPGDLGAPGSGPAGEP 624

Db 894 GSKPRGTGAGRGEPGAPGPPGPKGSGADGPIGAPGTPGQGIAGRGVVG 953

Qy 625 GPPGRRGPPGSGPAGDGVAGPKGAPGERGSGPAGP---KGSFGPAGPGEAGL 681

Db 954 GQRGPPGLPGPSGEPKQPSGASGERPPGPMGPPGLAGPGEAGREGAG 1013

Qy 682 GLTSGPSPGDPGKTGPPGAGQDORPPGPPGARGQAGVMGPPGPKGAAGEP 741

Db 1014 CRDGAAGPKGRGETGAPGPPGAPGAPGVPVPGAKNGDRGTGTGAPG 1073

Qy 742 GVPGPPGAVGPAKDGE-----AGAQGPP-----GPAGPAGER 774

Db 1074 GPAGPQSGRDKGTGEGDGMKGRHGFSLQPPPPGAGPQSGSGPAGP 1133

Qy 775 GEGPAGSPGQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPAG 820

Db 1134 GSAGAAKDLGLNGLPGIPGPRGRTGEVGVGPPGPPGPPGPP 1179

RESULT 11

OS 093251 PRELIMINARY; PRT; 1445 AA.

AC 093251; DT: 01-NOV-1998 (TReMBLrel. 08, Created)

DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)

DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)

DE Alpha 1 type I collagen.

GN Name=alpha 1 type I collagen;

OS Rana catesbeiana (Bull frog);

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Rana;

OX NCBI\_TaxID=8400;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9294154; PubMed=10367734;

RA Asahina K., Utoch R., Osawa M., Yoshizato K.;

RT "Cell-type specific and thyroid hormone-dependent expression of genes

RT of alpha 1(I) and alpha 2(I) collagen in intestine during

RT amphibian metamorphosis";

RL Matrix Biol. 18:89-103(1999).

DR EMBL; AB015440; BAA29028.1; --

DR GO; GO:0005581; C:cytoplasm; IEA.

DR GO; GO:0005737; C:cytoplasm; IEA.

DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.

DR GO; GO:0006817; P:phosphate transport; IEA.

DR InterPro; IPR008161; Clg helix.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR000885; Fib collagen\_C.

DR InterPro; IPR001007; VWF\_C.

DR Pfam; PF01410; COLFI; 1.

DR Pfam; PF01391; Collagen; 18.

DR ProDom; PD000007; Clg\_helix; 4.

DR ProDom; PD002078; Fib\_collagen\_C; 1.

DR SMART; SM00038; COLFI; 1.

DR SMART; SM00214; VWF; 1.

DR PROSITE; PS01208; VWF\_C\_1; 1.

DR PROSITE; PS01208; VWF\_C\_2; 1.

DR PROSITE; PS01208; VWF\_C\_3; 1.

DR PROSITE; PS01208; VWF\_C\_4; 1.

DR PROSITE; PS01208; VWF\_C\_5; 1.

DR PROSITE; PS01208; VWF\_C\_6; 1.

DR PROSITE; PS01208; VWF\_C\_7; 1.

DR PROSITE; PS01208; VWF\_C\_8; 1.

DR PROSITE; PS01208; VWF\_C\_9; 1.

DR PROSITE; PS01208; VWF\_C\_10; 1.

DR PROSITE; PS01208; VWF\_C\_11; 1.

DR PROSITE; PS01208; VWF\_C\_12; 1.

DR PROSITE; PS01208; VWF\_C\_13; 1.

DR PROSITE; PS01208; VWF\_C\_14; 1.

DR PROSITE; PS01208; VWF\_C\_15; 1.

DR PROSITE; PS01208; VWF\_C\_16; 1.

DR PROSITE; PS01208; VWF\_C\_17; 1.

DR PROSITE; PS01208; VWF\_C\_18; 1.

DR PROSITE; PS01208; VWF\_C\_19; 1.

DR PROSITE; PS01208; VWF\_C\_20; 1.

DR PROSITE; PS01208; VWF\_C\_21; 1.

DR PROSITE; PS01208; VWF\_C\_22; 1.

DR PROSITE; PS01208; VWF\_C\_23; 1.

DR PROSITE; PS01208; VWF\_C\_24; 1.

DR PROSITE; PS01208; VWF\_C\_25; 1.

DR PROSITE; PS01208; VWF\_C\_26; 1.

DR PROSITE; PS01208; VWF\_C\_27; 1.

DR PROSITE; PS01208; VWF\_C\_28; 1.

DR PROSITE; PS01208; VWF\_C\_29; 1.

DR PROSITE; PS01208; VWF\_C\_30; 1.

DR PROSITE; PS01208; VWF\_C\_31; 1.

DR PROSITE; PS01208; VWF\_C\_32; 1.

DR PROSITE; PS01208; VWF\_C\_33; 1.

DR PROSITE; PS01208; VWF\_C\_34; 1.

DR PROSITE; PS01208; VWF\_C\_35; 1.

DR PROSITE; PS01208; VWF\_C\_36; 1.

DR PROSITE; PS01208; VWF\_C\_37; 1.

DR PROSITE; PS01208; VWF\_C\_38; 1.

DR PROSITE; PS01208; VWF\_C\_39; 1.

DR PROSITE; PS01208; VWF\_C\_40; 1.

DR PROSITE; PS01208; VWF\_C\_41; 1.

DR PROSITE; PS01208; VWF\_C\_42; 1.

DR PROSITE; PS01208; VWF\_C\_43; 1.

DR PROSITE; PS01208; VWF\_C\_44; 1.

DR PROSITE; PS01208; VWF\_C\_45; 1.

DR PROSITE; PS01208; VWF\_C\_46; 1.

DR PROSITE; PS01208; VWF\_C\_47; 1.

DR PROSITE; PS01208; VWF\_C\_48; 1.

DR PROSITE; PS01208; VWF\_C\_49; 1.

DR PROSITE; PS01208; VWF\_C\_50; 1.

DR PROSITE; PS01208; VWF\_C\_51; 1.

DR PROSITE; PS01208; VWF\_C\_52; 1.

DR PROSITE; PS01208; VWF\_C\_53; 1.

DR PROSITE; PS01208; VWF\_C\_54; 1.

DR PROSITE; PS01208; VWF\_C\_55; 1.

DR PROSITE; PS01208; VWF\_C\_56; 1.

DR PROSITE; PS01208; VWF\_C\_57; 1.

DR PROSITE; PS01208; VWF\_C\_58; 1.

DR PROSITE; PS01208; VWF\_C\_59; 1.

DR PROSITE; PS01208; VWF\_C\_60; 1.

DR PROSITE; PS01208; VWF\_C\_61; 1.

DR PROSITE; PS01208; VWF\_C\_62; 1.

DR PROSITE; PS01208; VWF\_C\_63; 1.

DR PROSITE; PS01208; VWF\_C\_64; 1.

DR PROSITE; PS01208; VWF\_C\_65; 1.

DR PROSITE; PS01208; VWF\_C\_66; 1.

DR PROSITE; PS01208; VWF\_C\_67; 1.

DR PROSITE; PS01208; VWF\_C\_68; 1.

DR PROSITE; PS01208; VWF\_C\_69; 1.

DR PROSITE; PS01208; VWF\_C\_70; 1.

DR PROSITE; PS01208; VWF\_C\_71; 1.

DR PROSITE; PS01208; VWF\_C\_72; 1.

DR PROSITE; PS01208; VWF\_C\_73; 1.

DR PROSITE; PS01208; VWF\_C\_74; 1.

DR PROSITE; PS01208; VWF\_C\_75; 1.

DR PROSITE; PS01208; VWF\_C\_76; 1.

DR PROSITE; PS01208; VWF\_C\_77; 1.

DR PROSITE; PS01208; VWF\_C\_78; 1.

DR PROSITE; PS01208; VWF\_C\_79; 1.

DR PROSITE; PS01208; VWF\_C\_80; 1.

DR PROSITE; PS01208; VWF\_C\_81; 1.

DR PROSITE; PS01208; VWF\_C\_82; 1.

DR PROSITE; PS01208; VWF\_C\_83; 1.

DR PROSITE; PS01208; VWF\_C\_84; 1.

DR PROSITE; PS01208; VWF\_C\_85; 1.

DR PROSITE; PS01208; VWF\_C\_86; 1.

DR PROSITE; PS01208; VWF\_C\_87; 1.

DR PROSITE; PS01208; VWF\_C\_88; 1.

DR PROSITE; PS01208; VWF\_C\_89; 1.

DR PROSITE; PS01208; VWF\_C\_90; 1.

DR PROSITE; PS01208; VWF\_C\_91; 1.

DR PROSITE; PS01208; VWF\_C\_92; 1.

DR PROSITE; PS01208; VWF\_C\_93; 1.

DR PROSITE; PS01208; VWF\_C\_94; 1.

DR PROSITE; PS01208; VWF\_C\_95; 1.

DR PROSITE; PS01208; VWF\_C\_96; 1.

DR PROSITE; PS01208; VWF\_C\_97; 1.

DR PROSITE; PS01208; VWF\_C\_98; 1.

DR PROSITE; PS01208; VWF\_C\_99; 1.

DR PROSITE; PS01208; VWF\_C\_100; 1.

DR PROSITE; PS01208; VWF\_C\_101; 1.

DR PROSITE; PS01208; VWF\_C\_102; 1.

DR PROSITE; PS01208; VWF\_C\_103; 1.

DR PROSITE; PS01208; VWF\_C\_104; 1.

DR PROSITE; PS01208; VWF\_C\_105; 1.

DR PROSITE; PS01208; VWF\_C\_106; 1.

DR PROSITE; PS01208; VWF\_C\_107; 1.

DR PROSITE; PS01208; VWF\_C\_108; 1.

DR PROSITE; PS01208; VWF\_C\_109; 1.

DR PROSITE; PS01208; VWF\_C\_110; 1.

DR PROSITE; PS01208; VWF\_C\_111; 1.

DR PROSITE; PS01208; VWF\_C\_112; 1.

DR PROSITE; PS01208; VWF\_C\_113; 1.

DR PROSITE; PS01208; VWF\_C\_114; 1.

DR PROSITE; PS01208; VWF\_C\_115; 1.

DR PROSITE; PS01208; VWF\_C\_116; 1.

DR PROSITE; PS01208; VWF\_C\_117; 1.

DR PROSITE; PS01208; VWF\_C\_118; 1.

DR PROSITE; PS01208; VWF\_C\_119; 1.

DR PROSITE; PS01208; VWF\_C\_120; 1.

DR PROSITE; PS01208; VWF\_C\_121; 1.

DR PROSITE; PS01208; VWF\_C\_122; 1.

DR PROSITE; PS01208; VWF\_C\_123; 1.

DR PROSITE; PS01208; VWF\_C\_124; 1.

DR PROSITE; PS01208; VWF\_C\_125; 1.

DR PROSITE; PS01208; VWF\_C\_126; 1.

DR PROSITE; PS01208; VWF\_C\_127; 1.

DR PROSITE; PS01208; VWF\_C\_128; 1.

DR PROSITE; PS01208; VWF\_C\_129; 1.

DR PROSITE; PS01208; VWF\_C\_130; 1.

DR PROSITE; PS01208; VWF\_C\_131; 1.

DR PROSITE; PS01208; VWF\_C\_132; 1.

DR PROSITE; PS01208; VWF\_C\_133; 1.

DR PROSITE; PS01208; VWF\_C\_134; 1.

DR PROSITE; PS01208; VWF\_C\_135; 1.

DR PROSITE; PS01208; VWF\_C\_136; 1.

DR PROSITE; PS01208; VWF\_C\_137; 1.

DR PROSITE; PS01208; VWF\_C\_138; 1.

DR PROSITE; PS01208; VWF\_C\_139; 1.

DR PROSITE; PS01208; VWF\_C\_140; 1.

DR PROSITE; PS01208; VWF\_C\_141; 1.

DR PROSITE; PS01208; VWF\_C\_142; 1.

DR PROSITE; PS01208; VWF\_C\_143; 1.

DR PROSITE; PS01208; VWF\_C\_144; 1.

DR PROSITE; PS01208; VWF\_C\_145; 1.

DR PROSITE; PS01208; VWF\_C\_146; 1.

DR PROSITE; PS01208; VWF\_C\_147; 1.

DR PROSITE; PS01208; VWF\_C\_148; 1.

DR PROSITE; PS01208; VWF\_C\_149; 1.

DR PROSITE; PS01208; VWF\_C\_150; 1.

DR PROSITE; PS01208; VWF\_C\_151; 1.

DR PROSITE; PS01208; VWF\_C\_152; 1.

DR PROSITE; PS01208; VWF\_C\_153; 1.

DR PROSITE; PS01208; VWF\_C\_154; 1.

DR PROSITE; PS01208; VWF\_C\_155; 1.

DR PROSITE; PS01208; VWF\_C\_156; 1.

DR PROSITE; PS01208; VWF\_C\_157; 1.

DR PROSITE; PS01208; VWF\_C\_158; 1.

DR PROSITE; PS01208; VWF\_C\_159; 1.

DR PROSITE; PS01208; VWF\_C\_160; 1.

DR PROSITE; PS01208; VWF\_C\_161; 1.

DR PROSITE; PS01208; VWF\_C\_162; 1.

DR PROSITE; PS01208; VWF\_C\_163; 1.

DR PROSITE; PS01208; VWF\_C\_164; 1.

DR PROSITE; PS01208; VWF\_C\_165; 1.

DR PROSITE; PS01208; VWF\_C\_166; 1.

DR PROSITE; PS01208; VWF\_C\_167; 1.

DR PROSITE; PS01208; VWF\_C\_168; 1.

DR PROSITE; PS01208; VWF\_C\_169; 1.

DR PROSITE; PS01208; VWF\_C\_170; 1.

DR PROSITE; PS01208; VWF\_C\_171; 1.

DR PROSITE; PS01208; VWF\_C\_172; 1.

DR PROSITE; PS01208; VWF\_C\_173; 1.

DR PROSITE; PS01208; VWF\_C\_174; 1.

DR PROSITE; PS01208; VWF\_C\_175; 1.

DR PROSITE; PS01208; VWF\_C\_176; 1.

DR PROSITE; PS01208; VWF\_C\_177; 1.

DR PROSITE; PS01208; VWF\_C\_178; 1.

DR PROSITE; PS01208; VWF\_C\_179; 1.

DR PROSITE; PS01208; VWF\_C\_180; 1.

DR PROSITE; PS01208; VWF\_C\_181; 1.

DR PROSITE; PS01208; VWF\_C\_182; 1.

DR PROSITE; PS01208; VWF\_C\_183; 1.

DR PROSITE; PS01208; VWF\_C\_184; 1.

DR PROSITE; PS01208; VWF\_C\_185; 1.

DR PROSITE; PS01208; VWF\_C\_186; 1.

DR PROSITE; PS01208; VWF\_C\_187; 1.

DR PROSITE; PS01208; VWF\_C\_188; 1.

DR PROSITE; PS01208; VWF\_C\_189; 1.

DR PROSITE; PS01208; VWF\_C\_190; 1.

DR PROSITE; PS01208; VWF\_C\_191; 1.

DR PROSITE; PS01208; VWF\_C\_192; 1.

DR PROSITE; PS01208; VWF\_C\_193; 1.

DR PROSITE; PS01208; VWF\_C\_194; 1.

DR PROSITE; PS01208; VWF\_C\_195; 1.

DR PROSITE; PS01208; VWF\_C\_196; 1.

DR PROSITE; PS01208; VWF\_C\_197; 1.

DR PROSITE; PS01208; VWF\_C\_198; 1.

DR PROSITE; PS01208; VWF\_C\_199; 1.

DR PROSITE; PS01208; VWF\_C\_200; 1.

DR PROSITE; PS01208; VWF\_C\_201; 1.

DR PROSITE; PS01208; VWF\_C\_202; 1.

DR PROSITE; PS01208; VWF\_C\_203; 1.

DR PROSITE; PS01208; VWF\_C\_204; 1.

DR PROSITE; PS01208; VWF\_C\_205; 1.

DR PROSITE; PS01208; VWF\_C\_206; 1.

DR PROSITE; PS01208; VWF\_C\_207; 1.

DR PROSITE; PS01208; VWF\_C\_208; 1.

DR PROSITE; PS01208; VWF\_C\_209; 1.

DR PROSITE; PS01208; VWF\_C\_210; 1.

DR PROSITE; PS01208; VWF\_C\_211; 1.

DR PROSITE; PS01208; VWF\_C\_212; 1.

DR PROSITE; PS01208; VWF\_C\_213; 1.

DR PROSITE; PS01208; VWF\_C\_214; 1.

DR PROSITE; PS01208; VWF\_C\_215; 1.

DR PROSITE; PS01208; VWF\_C\_216; 1.

DR PROSITE; PS01208; VWF\_C\_217; 1.

DR PROSITE; PS01208; VWF\_C\_218; 1.

DR PROSITE; PS01208; VWF\_C\_219; 1.

DR PROSITE; PS01208; VWF\_C\_220; 1.

DR PROSITE; PS01208; VWF\_C\_221; 1.

DR PROSITE; PS01208; VWF\_C\_222; 1.

DR PROSITE; PS01208; VWF\_C\_223; 1.

DR PROSITE; PS01208; VWF\_C\_224; 1.

DR PROSITE; PS01208; VWF\_C\_225; 1.

DR PROSITE; PS01208; VWF\_C\_226; 1.

DR PROSITE; PS01208; VWF\_C\_227; 1.

DR PROSITE; PS01208; VWF\_C\_228; 1.

DR PROSITE; PS01208; VWF\_C\_229; 1.

DR PROSITE; PS01208; VWF\_C\_230; 1.

DR PROSITE; PS01208; VWF\_C\_231; 1.

DR PROSITE; PS01208; VWF\_C\_232; 1.

DR PROSITE; PS01208; VWF\_C\_233; 1.

DR PROSITE; PS01208; VWF\_C\_234; 1.

DR PROSITE; PS01208; VWF\_C\_235; 1.

DR PROSITE; PS01208; VWF\_C\_236; 1.

DR PROSITE; PS01208; VWF\_C\_237; 1.

DR PROSITE; PS01208; VWF\_C\_238; 1.

DR PROSITE; PS01208; VWF\_C\_239; 1.

DR PROSITE; PS01208; VWF\_C\_240; 1.

DR PROSITE; PS01208; VWF\_C\_241; 1.

DR PROSITE; PS01208; VWF\_C\_242; 1.

DR PROSITE; PS01208; VWF\_C\_243; 1.

DR PROSITE; PS01208; VWF\_C\_244; 1.

DR PROSITE; PS01208; VWF\_C\_245; 1.

DR PROSITE; PS01208; VWF\_C\_246; 1.

DR PROSITE; PS01208; VWF\_C\_247; 1.

DR PROSITE; PS01208; VWF\_C\_248; 1.

DR PROSITE; PS01208; VWF\_C\_249; 1.

DR PROSITE; PS01208; VWF\_C\_250; 1.

DR PROSITE; PS01208; VWF\_C\_251; 1.

DR PROSITE; PS01208; VWF\_C\_252; 1.

DR PROSITE; PS01208; VWF\_C\_253; 1.

DR PROSITE; PS01208; VWF\_C\_254; 1.

DR PROSITE; PS01208; VWF\_C\_255; 1.

DR PROSITE; PS01208; VWF\_C\_256; 1.

DR PROSITE; PS01208; VWF\_C\_257; 1.

DR PROSITE; PS01208; VWF\_C\_258; 1.

DR PROSITE; PS01208; VWF\_C\_259; 1.

DR PROSITE; PS01208; VWF\_C\_260; 1.

DR PROSITE; PS01208; VWF\_C\_261; 1.

DR PROSITE; PS01208; VWF\_C\_262; 1.

DR PROSITE; PS01208; VWF\_C\_263; 1.

DR PROSITE; PS01208; VWF\_C\_264; 1.

DR PROSITE; PS01208; VWF\_C\_265; 1.

DR PROSITE; PS01208; VWF\_C\_266; 1.

DR PROSITE; PS01208; VWF\_C\_267; 1.

DR PROSITE; PS01208; VWF\_C\_268; 1.

DR PROSITE; PS01208; VWF\_C\_269; 1.

DR PROSITE; PS01208; VWF\_C\_270; 1.

DR PROSITE; PS01208; VWF\_C\_271; 1.

DR PROSITE; PS01208; VWF\_C\_272; 1.

DR PROSITE; PS01208; VWF\_C\_273; 1.

DR PROSITE; PS01208; VWF\_C\_274; 1.

DR PROSITE; PS01208; VWF\_C\_275; 1.

DR PROSITE; PS01208; VWF\_C\_276; 1.

DR PROSITE; PS01208; VWF\_C\_277; 1.

DR PROSITE; PS01208; VWF\_C\_278; 1.

DR PROSITE; PS01208; VWF\_C\_279; 1.

DR PROSITE; PS01208; VWF\_C\_280; 1.

DR PROSITE; PS01208; VWF\_C\_281; 1.

DR PROSITE; PS01208; VWF\_C\_282; 1.

DR PROSITE; PS01208; VWF\_C\_283; 1.

DR PROSITE; PS01208; VWF\_C\_284; 1.

DR PROSITE; PS01208; VWF\_C\_285; 1.

DR PROSITE; PS01208; VWF\_C\_286; 1.

DR PROSITE; PS01208; VWF\_C\_287; 1.

DR PROSITE; PS01208; VWF\_C\_288; 1.

DR PROSITE; PS01208; VWF\_C\_289; 1.

DR PROSITE; PS01208; VWF\_C\_290; 1.

DR PROSITE; PS01208; VWF\_C\_291; 1.

DR PROSITE; PS01208; VWF\_C\_292; 1.

DR PROSITE; PS01208; VWF\_C\_293; 1.

DR PROSITE; PS01208; VWF\_C\_294; 1.

DR PROSITE; PS01208; VWF\_C\_295; 1.

DR PROSITE; PS01208; VWF\_C\_296; 1.

DR PROSITE; PS01208; VWF\_C\_297; 1.

DR PROSITE; PS01208; VWF\_C\_298; 1.

DR PROSITE; PS01208; VWF\_C\_299; 1.

DR PROSITE; PS01208; VWF\_C\_300; 1.

DR PROSITE; PS01208; VWF\_C\_301; 1.

DR PROSITE; PS01208; VWF\_C\_302; 1.

DR PROSITE; PS01208; VWF\_C\_303; 1.

DR PROSITE; PS01208; VWF\_C\_304; 1.

DR PROSITE; PS01208; VWF\_C\_305; 1.

DR PROSITE; PS01208; VWF\_C\_306; 1.

DR PROSITE; PS01208; VWF\_C\_307; 1.

DR PROSITE; PS01208; VWF\_C\_308; 1.

DR PROSITE; PS01208; VWF\_C\_309; 1.

DR PROSITE; PS01208; VWF\_C\_310; 1.

DR PROSITE; PS01208; VWF\_C\_311; 1.

DR PROSITE; PS01208; VWF\_C\_312; 1.

DR PROSITE; PS01208; VWF\_C\_313; 1.

DR PROSITE; PS01208; VWF\_C\_314; 1.

DR PROSITE; PS01208; VWF\_C\_315; 1.

DR PROSITE; PS01208; VWF\_C\_316; 1.

DR PROSITE; PS01208; VWF\_C\_317; 1.

DR PROSITE; PS01208; VWF\_C\_318; 1.

DR PROSITE; PS01208; VWF\_C\_319; 1.

DR PROSITE; PS01208; VWF\_C\_320; 1.

DR PROSITE; PS01208; VWF\_C\_321; 1.

DR PROSITE; PS01208; VWF\_C\_322; 1.

DR PROSITE; PS01208; VWF\_C\_323; 1.

DR PROSITE; PS01208; VWF\_C\_324; 1.

DR PROSITE; PS01208; VWF\_C\_325; 1.

DR PROSITE; PS01208; VWF\_C\_326; 1.

DR PROSITE; PS01208; VWF\_C\_327; 1.

DR PROSITE; PS01208; VWF\_C\_328; 1.

DR PROSITE; PS01208; VWF\_C\_329; 1.

DR PROSITE; PS01208; VWF\_C\_330; 1.

DR PROSITE; PS01208; VWF\_C\_331; 1.

DR PROSITE; PS01208; VWF\_C\_332; 1.

DR PROSITE; PS01208; VWF\_C\_333; 1.

DR PROSITE; PS01208; VWF\_C\_334; 1.

DR PROSITE; PS01208; VWF\_C\_335; 1.

DR PROSITE; PS01208; VWF\_C\_336; 1.

DR PROSITE; PS01208; VWF\_C\_337; 1.

DR PROSITE; PS01208; VWF\_C\_338; 1.

DR PROSITE; PS01208; VWF\_C\_339; 1.

DR PROSITE; PS01208; VWF\_C\_340; 1.

DR PROSITE; PS01208; VWF\_C\_341; 1.

DR PROSITE; PS01208; VWF\_C\_342; 1.

DR PROSITE; PS01208; VWF\_C\_343; 1.

DR PROSITE; PS01208; VWF\_C\_344; 1.

DR PROSITE; PS01208; VWF\_C\_345; 1.

DR PROSITE; PS01208; VWF\_C\_346; 1.

DR PROSITE; PS01208; VWF\_C\_347; 1.

DR PROSITE; PS01208; VWF\_C\_348; 1.

DR PROSITE; PS01208; VWF\_C\_349; 1.

DR PROSITE; PS01208; VWF\_C\_350; 1.

DR PROSITE; PS01208; VWF\_C\_351; 1.

DR PROSITE; PS01208; VWF\_C\_352; 1.

DR PROSITE; PS01208; VWF\_C\_353; 1.

DR PROSITE; PS01208; VWF\_C\_354; 1.

DR PROSITE; PS01208; VWF\_C\_355; 1.

DR PROSITE; PS01208; VWF\_C\_356; 1.

DR PROSITE; PS01208; VWF\_C\_357; 1.

DR PROSITE; PS01208; VWF\_C\_358; 1.

DR PROSITE; PS01208; VWF\_C\_359; 1.

DR PROSITE; PS01208; VWF\_C\_360; 1.

DR PROSITE; PS01208; VWF\_C\_361; 1.

DR PROSITE; PS01208; VWF\_C\_362; 1.

DR PROSITE; PS01208; VWF\_C\_363; 1.

DR PROSITE; PS01208; VWF\_C\_364; 1.

DR PROSITE; PS01208; VWF\_C\_365; 1.

DR PROSITE; PS01208; VWF\_C\_366; 1.

DR PROSITE; PS01208; VWF\_C\_367; 1.

DR PROSITE; PS01208; VWF\_C\_368; 1.

DR PROSITE; PS01208; VWF\_C\_369; 1.

DR PROSITE; PS01208; VWF\_C\_370; 1.

DR PROSITE; PS01208; VWF\_C\_371; 1.

DR PROSITE; PS01208; VWF\_C\_372; 1.

DR PROSITE; PS01208; VWF\_C\_373; 1.

DR PROSITE; PS01208; VWF\_C\_374; 1.

DR PROSITE; PS01208; VWF\_C\_375; 1.

DR PROSITE; PS01208; VWF\_C\_376; 1.

DR PROSITE; PS01208; VWF\_C\_377; 1.

DR PROSITE; PS01208; VWF\_C\_378; 1.

DR PROSITE; PS01208; VWF\_C\_379; 1.

DR PROSITE; PS01208; VWF\_C\_380; 1.

DR PROSITE; PS01208; VWF\_C\_381; 1.

DR PROSITE; PS01208; VWF\_C\_382; 1.

DR PROSITE; PS01208; VWF\_C\_383; 1.

DR PROSITE; PS01208; VWF\_C\_384; 1.

DR PROSITE; PS01208; VWF\_C\_385; 1.

DR PROSITE; PS01208; VWF\_C\_386; 1.

DR PROSITE; PS01208; VWF\_C\_387; 1.

DR PROSITE; PS01208; VWF\_C\_388; 1.

DR PROSITE; PS01208; VWF\_C\_389; 1.

DR PROSITE; PS01208; VWF\_C\_390; 1.

DR PROSITE; PS01208; VWF\_C\_391; 1.

DR PROSITE; PS01208; VWF\_C\_392; 1.

DR PROSITE; PS01208; VWF\_C

```

Db 1061 GPAGPAGARGSGAGARGDKGAGEQGERGMKCHRGFNLDLPDPGAPGAGHAGEGPGSGAS 1120
QY 763 GPPGPAGPAGRGQGPAGSGFQGLPFPAGPPGAGKPGSE---QGVPGDLGAPGPGSPA 819
Db 1121 GPAGPRGPGSGSGPGKDGSG---NGLPQIPGPPGRKRTGVDVGPAGPPGPPGPP 1177
QY 820 GG 821
Db 1178 GG 1179

RESULT 12
QYIB4 PRELIMINARY; PRT; 1450 AA.
AC Q9YIB4
AC Q9YIB4
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen.
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX NCBI_TaxID=8330;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Regenerate forelimbs;
RX MEDLINE=99407244; PubMed=1047166;
RA Asahina K., Obara M., Yoshizato K.;
RT "Expression of genes of type I and type II collagen in the formation
RL and development of the blastema of regenerating newt limb.";
RL Dev. Dyn. 216:59-71(1999).
DR EMBL; AB015438; BAA36973.1;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; E:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLF1; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; C1g_helix; 4.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLF1; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; UNKNOWN_1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen.
SQ SEQUENCE 1450 AA; 137563 MW; ABF8A74841B87B7C CRC64;

Query Match 55.2%; Score 2561.5; DB 2; Length 1450;
Best Local Similarity 52.3%; Pred. No. 3.4e-95;
Matches 517; Conservative 39; Mismatches 261; Indels 171; Gaps 17;

QY 1 GPPGEP-----GPTGLGPPGPERG-----GP----- 21
Db 192 GEFPGAAGALGRLGLPFGPKNGDDGESKPGRPGRGSGPQAGRLFTAGLPQMK 251
QY 22 GSRGFPAGDVAGPKGPA-----GERSPGAPGKPGSGPAGRPGEAGLPLCAKGLT 72
Db 252 GHRGNGLDKAGDNGPAGPKPEPKNPENGAFGQAGPGLFGERGRGAGPGAGRGND 311
QY 73 GSPGSGPDPGKTPGPP-----AGDGRGPPGPPGARGQAGVNGFPGPKGAA 120
Db 312 GSPGAAGPPGPTGPTGPPGPGAVGAKGDAGPQSGRSGEPQAGRGEPA---FPAAGAA 368
QY 121 GEPKXAGEGVGPPGAVP---AKDGEAGAQGPAGPAGGERGEGPAGSGFGGLP 177
Db 369 GFSGNPPTDQPGGKGATGSPGIAGAPFPFGARGAPFGQAGAPGPKNGNGEPAQNK 428
QY 178 GPAGPPGAGKPGEGQVPGDLGAPGSPGAPGPGTGLPFGPPGGRGSGRFGGADGVA 237

```

```

Db 429 GEPGAKGERPGAVGQVPPSPGSEEGKRGSRGEPGPPGAPGGERGSGRGPFGSDGAS 488
QY 238 GPKXGAPAGERGSPGAPGKSGPGEAGRPGHAGLPGAKGLTGSGSGSPGDPDKTGPDPAGOD 297
Db 489 GPKGAPGERGSGVGPAGPKSGTGSRPGEPGGLPGAKGLTGSGSGSPGDPDKTGPAGAAGOD 548
QY 298 GRPPEGPPGARGQAGVNGFPGPKGAAGBPGKAGRGVPPGPAVGPAGKDGEGAGAQGPP 357
Db 549 GHPGPPGSPGARGQAGVNGFPGPKGAAGBPGKAGRGVPPGPAVGPAGKDGEGAGAQGPP 608
QY 358 GPAGPAGERGEGQAGSPGFGQLPFPAGPPGAGKPGGEGVPGDLGAPGSPGAPGE--- 413
Db 609 GFSGSPGERGEGQAGSPGFGQLPFPAGPPGAGKPGGEGVPGDLGAPGSPGAPGE--- 668
QY 414 -----PGPTGLGSP-----PGERGGPGSGRGP----- 435
Db 669 GERGGQGPAGAGPRGSPGSPGNDGAKGAGAGAPGGRGPPGLQMPGERSAGMPGAK 728
QY 436 -----GADGVAGPKGPAGRGSP-----GPAGPKSGPGEAGRP 468
Db 729 GDRGDAGTKGADGAPGKDGARGLTGPIGPPGSGAPGDKGEGGSPGAPGTGARGSPGR 788
QY 469 GEAGLPGAKGLTGSPGS-----PGPDGKTGPPGAGQDGRPPGPPGARGQAGVNGFPGPK 525
Db 789 GEPGAPGAPGICGPPGADGQPGAKGESGDAGPKDAGAFGAPGTGAPGAGNVGAPGPK 848
QY 526 GAAGEPKAGRGVPPGPPGAVGAPGAKDGAGAQGPPGPPGAPGAGERGEGQAGSPGFGGLP 585
Db 849 GTRGAGAPGATGFFCAAGELGPPGSPGNAGPPGPPGPGKSGKSGRGETGPAGRSRP 908
QY 586 GPAGPP---GEAKPGEGVPGDLGAPGSGPAGBPGPTGLPQPPQERGGSGRSGPPGAD 542
Db 909 GPAGPPGSPGKSGSGSDGAPAGPIGPGQIAGRGVVLPGQGRGERGFSGLPGSPGPP 968
QY 643 GVAGPKPAGERGSPGAPG-----KGSPEAGRPGEAGLPGAK-----GLT 684
Db 969 GKQPSGPNGERGPPGPPGPPGLGPPGREGSGSGAGPRDGSPPGKDRGENGSP 1028
QY 685 GSPGSGPDPKTPGPPGAPGD---GRPPGPPGPPGARGQAGVNGFPGPKGAAGBPGKAGE- 740
Db 1029 GPPGAPGAPGAPGVPVGPAGKNGDRGTGPAGPAGPAGSGVAGPAGPAGARGDKGAGEQ 1088
QY 741 -----RGVGPVAVGAPGAKDGAGAQGPPGPPGAPGAGERGEGQAGSPGFGGLP 792
Db 1089 GERGMKGRHGFNGMQPPGPPGSSGEGQAGSGPAGPRGPPGSGSGTGRDGVNGLPPI 1148
QY 793 GPPGKAGKPGEGQVPGDLGAPGSPGAPG 820
Db 1149 GPPGPRGNDVGPAGPPGPPGPPG 1176

RESULT 13
Q802B5 PRELIMINARY; PRT; 1449 AA.
AC Q802B5
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cellial-prov protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Embryo;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

```

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzyzanski M.L., Skalska U., Smallus D.E., Schmerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RA "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=22341132; PubMed=12454917;  
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
 RA Richardson P.;  
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 RT initiative.";  
 RL Dev. Dyn. 225:384-391(2002).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Klein S., Strausberg R.;  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC049829; AAH49829.1; -;  
 DR GO; GO:0005581; C:collagen; IEA.  
 DR GO; GO:0005737; C:cytoplasm; IEA.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR GO; GO:0006817; P:phosphate transport; IEA.  
 DR InterPro; IPR008161; C1g helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR000885; Fib collagen\_C.  
 DR InterPro; IPR001007; VWF\_C.  
 DR Pfam; PF01410; COLF1; 1.  
 DR Pfam; PF01391; Collagen; 18.  
 DR Pfam; PF00093; VWC; 1.  
 DR ProDom; PD000007; C1g helix; 4.  
 DR ProDom; PD002078; Fib collagen\_C; 1.  
 DR SMART; SM00038; COLF1; 1.  
 DR SMART; SM00214; VWC; 1.  
 DR PROSITE; PS01208; VWF\_C.1; UNKNOWN 1.  
 DR PROSITE; PS0184; VWF\_C.2; 1.  
 KW Collagen.  
 SQ SEQUENCE 1449 AA; 137464 MW; 4B32E20BCECE978C CRC64;

Query Match 54.5%; Score 2529; DB 2; Length 1449;  
 Best Local Similarity 54.6%; Pred No. 57e-94;  
 Matches 500; Conservative 53; Mismatches 261; Indels 102; Gaps 13;

QY 1 GPPGPGTGLPPTGPGGPGS-----RGPGADGVAGPKGPA-----GER 42  
 Db 222 CKPRGPRGPRGPPGQARGLPCTAGLPCKMGRHGNGLDGAAGSGPAGPKGPGSPGPN 281  
 QY 43 GSPGPGKSPGAGRGEAGLPKAGLTSGPSGPGDKTGPPT-----A 90  
 Db 282 GAGQVGRGSGRGRPPGPGARGNDGAPGAGPSTGSPGPPGPGVGPKGDA 341

QY 91 QDGRGPPGPPGARGQAGVGMFPKPGAGGEPKAGRGVPPGPPGAVGPAKDGAGAG 150  
 Db 342 GPGSGSDGPGQGRGEPGAPGQAGAGSPGSGNPGSDGPKAGKATGATGAPGPPGAR 401

QY 151 GPPGPGAGRGPGGPGAGSGPGLPGPAGPPGAGKPGEGQVPGDLGAPGSGPAGEP 210  
 Db 402 GAPAGQGGGPGPKGNNGEPGACNKGAGKGEPPAGVQGPVPPGSGEGKRGSGEP 461

QY 211 GPTGLPGRGPGGPGGPPGADGVAGPKGAGERSGPPAGPKGSPGAGRPPGAGLP 270  
 Db 462 GPAGPPGAGRGGPGGPPGSDGAGPKGAGRGVPGAPKSGSGSGRPPGGLP 521

QY 271 GAKLTGTPSGSPGDPGKTGPPGAPGQDGRPRGPPGARGQAGVGMFPKPGAGGPGKA 330  
 Db 522 GAKLTGTPSGSPGSDGKTGTPAGAPQDGRAGPPGARGQAGVGMFPKPGAGGPGKN 581

QY 331 GERGVPPGAVGPPAGKGEAGAAQPPGPPGAPGARGGPGGPPGAGPPGPPGPPG 390  
 Db 582 GERGVAGPPGAVGLPKGQDAGAQPPGPPGAPGARGGPGGPPGAGPPGPPGPPG 641

QY 391 GKGEGGVPGDLGAGPSPAGE---PQPTGLPGRGPPGPPGPPGPPGPPGPPGPPG 447  
 Db 642 GRPGEGVPGDVPSPGAPARGRGPPGARGGPPGPPGPPGPPGPPGPPGPPGPPG 701

QY 448 GERGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 501  
 Db 702 GAPGGGPPGLQMPGER---GSSGLPKAGKDRGQGVKSGDTPGKDVRLTGP 757

QY 502 GRPGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 558  
 Db 758 --PGPGAPGDKGEGAPGAPGAGTGAAGPPGPPGPPGPPGPPGPPGPPGPPG 815

QY 559 -----GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 587  
 Db 815 GDSGAKGDAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 875

QY 588 --AGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 645  
 Db 876 GNAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 935

QY 646 GPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 702  
 Db 936 GPGVAGSGRTVGLPGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 995

QY 703 GQDGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 762  
 Db 996 GSGREGAPGSGCAPGRCGAVGPKGDRGEGGPPGPPGPPGPPGPPGPPGPPGPPG 1055

QY 763 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 804  
 Db 1056 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1115

QY 805 GVEGDLGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 820  
 Db 1116 GPSGASGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1131

RESULT 14  
 Q6PCL3  
 ID Q6PCL3 PRELIMINARY; PRT; 1225 AA.  
 AC Q6PCL3;  
 DT 05-JUL-2004 (TREMELrel. 27, Created)  
 DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)  
 DE Colial protein.  
 GN Name=Colial;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mix FVB/N;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Hopkins R.P., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusik M.B., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,



RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RL and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=Mix FVB/N;  
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;  
RC Strausberg R.; to the EMBL/GenBank/DBJ databases.  
RL Submitted (Oct-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC059281; AAH59281.1; -  
DR InterPro; IPR008161; Cig helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR002181; Fibrinogen C.  
DR InterPro; IPR000885; Fib collagen\_C.  
DR InterPro; IPR003041; PNP\_SGCI.  
DR InterPro; IPR001007; VWF\_C.  
DR Pfam; PF01410; COLF1; 1.  
DR Pfam; PF01391; Collagen; 14.  
DR Pfam; PF00093; VWC; 1.  
DR ProDom; PD000007; Cig helix; 2.  
DR ProDom; PD02078; Fib\_collagen\_C; 1.  
DR SMART; SM00038; COLF1; 1.  
DR SMART; SM00214; VWC; 1.  
DR PROSITE; PS01208; VWFC.1; 1.  
DR PROSITE; PS0184; VWFC.2; 1.  
KW Collagen.  
SQ SEQUENCE 1225 AA; 117860 MW; B6B86CB4457F4D9 CRC64;

Query Match 53.8%; Score 2494; DB 2; Length 1225;  
Best Local Similarity 56.0%; Pred. No. 1.5e-92;  
Matches 510; Conservative 29; Mismatches 227; Indels 144; Gaps 20;

QY 12 PGPGERGG----PGSRGFPACDGVAGPKGAGERSGPGAGPKGSPGAGRPGEGAGLP 66  
DB 73 PNFQRECECCAFCEEYVNSDEV-VGEGPKDGPQGPQGPVGPVPPGRDGI PQQGLP 131  
QY 67 GAKGLTGPSPGPPDGK-----TGPPGAGQDGRGPRGPPGARGQA 108  
DB 132 GPPGHGPPPPGLGNGFASQMSYGYDEKSAVSVPGPMGSPGRGUPPGAGPQ--- 188  
QY 109 GVMGPPGPKGAAGBPKAGERGVPVPGAVGPAKDGAGAGAGQGP---GPAGPAGERGEQ 165  
DB 189 ----GFGQPPGPEBEGEGSGPVGPRGPPGPKXGDDGAGKPRGERGPPGQARGLP 245  
QY 166 GPAGSP-----GFQGLPFPAGPGEAKPGEGQVPGDLGAPSGSPGAGEGPTGLPQPP 219  
DB 246 GTAGLPGMKHGRGFSGLDGAAGDAGPAGPKGSPSGENGAFQMGPRGLPGERGFGPP 305  
QY 220 GERGFSGSRGPGAGDGVAGPKGAGERSGPGAGPKGSPGAGRPGEGAGLPCAAGLTSP 279  
DB 306 GT---AGAGKNDGAVGAAGPPGPT-----GTTGPPGPAVGAKEAGAPQAR----- 350  
QY 280 GSPGPDGKTGPPGPAQDGRPPGPPGARGAGVGMFPFGKGAAGPECKAGERGVPPGP 339  
DB 351 GSEGGQVTRGEPGPPGPAAGAGPAGNPGADGQ-----PGAKGANGAPGIA---GAPGFP 401  
QY 340 GAVGPAKDGEGAGAQPPGPPGAPAGERGEGQAGSPFGQLPGPAGPGEAKPGEQVPP 399  
DB 402 GARGP-----SGPQPSGPPGPKNSGE-----PGAPGNKGDGTGAKGEPGATGVQPP 449  
QY 400 GDLGAPGSPGAGEPPTGLPVPBGRGGPGRGFPFAGDGVAGPKGAGERSGPPGAPGK 459  
DB 450 GPAGEGKKGARGEPGSLPFPBGRGGPGRGFPFAGDGVAGPKGSPGERGAPGAPGK 509  
QY 460 GSPGEAGRPGEAGLPKAGKLTGSPGSPGPDGKTGTPGPPAGQDGRGPPGPPGARGQAGVM 519

DB 510 GSPGEAGRPGEAGLPKAGKLTGSPGSPGPDGKTGTPGPPAGQDGRGPPGPPGARGQAGVM 569  
QY 520 GPPGPKGAAGBPKAGERGVPVPGAVGPAKDGAGAGAQGPAGPAGERGEGQAGSP 579  
DB 570 GFFGPKGTAGBPKAGERGLSPGCAVGPACDGEAGAGQAGPAGPAGERGEGQAGSP 629  
QY 580 GFQGLPGPAGPGEAGKPGEGVPGDLGAPSPSPAGE-----PGP----- 620  
DB 630 GFQGLPGPAGPGEAGKPGEGVPGDLGAPSPSPAGE-----PGP----- 689  
QY 621 -----TGLPGPGERGGPSRGFPFAGDGVAGPKGAGERSGPPGAPGKSPGE- 668  
DB 690 GAFGNDGAKGDTGAPGAPGSPGAPLQMPGARGNAGLPGPKGDEGDAGPKGADGSPGKD 749  
QY 669 -----AGRPGEAGLPCAAGLTGSPGSPGPDGKTGTPGPPAGQDGRGPPGPPGARGQA 720  
DB 750 GARGLTGTPGPPGPAAGPDKGAGSPGPPGPTGARGAPGDRGEAGPAGPAGFAGPPGAP 809  
QY 721 GVMGPPGPKGAAGBPKAGERGVPVPGAVGPAKDGAGAGAQGPAGPAGERGE--- 776  
DB 810 GAGGAPGVPVGPAGKNGDRGTGTPGAPGAPGIPGARGGPPGQGRDKGTGEGQDRIK 869  
QY 777 -----QGPAAGSPGFQGLPQPA-----GPPGEGAGKPGE---QVPGDLGAPGP- 815  
DB 870 GHRGFSGLGPPGSPGSPGEGQAGSPGAGPAGPRGPGSAGSPCKDGLNGLPGIPGPPGPR 929  
QY 816 -----SGPAG 820  
DB 930 GRTGDSGPAG 939

RESULT 15  
AAH59281  
ID AAH59281 PRELIMINARY; PRT; 1225 AA.  
AC AAH59281;  
DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)  
DE Collal protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Mammary tumor;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Wang J., Wang J., Hsieh F.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RL and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Mammary tumor;  
RX Strausberg R.;  
RA Submitted (Oct-2003) to the EMBL/GenBank/DBJ databases.  
RL

DR EMBL: BC059281.1; --  
SQ SEQUENCE 1225 AA; 117860 MW; B6B86CBB4457F4D9 CRC64;

Query Match 53.8%; Score 2494; DB 2; Length 1225;  
Best Local Similarity 56.0%; Pred. No. 1.5e-92;  
Matches 510; Conservative 29; Mismatches 227; Indels 144; Gaps 20;

```
QY 12 PCPPGERGG-----PGRGFGGADGAGVAGPKGAGGSGPGPAGPKGSPGCEAGRFGCEAGLP 66
Db 73 PNPQRREGCCAFCEYYVSPNSDEV-GVEGPKGDPGPGPVPVGPVGPVGPVGPVGPVGPVGPVGP 131
QY 67 GAKGLTSGSPGSPGPDGK-----TGPPGAGODGRGPPGPPGARGQA 108
Db 132 GPFGHPPGPPGLGNFASQMSYCYDEKSAGSVFPGMGSGPRGLPGPCAPGQ--- 189
QY 109 GVMGFPKGAAGEFGKAGRGVPGPPGAVGAPGAKDGAAGAGGPP---GPAPAGERGEQ 165
Db 189 ---GFQPPGEPGEGGSGPMGPRGPPGPKNGDDGEAGKPRFGRGPPGPPGQARGLP 245
QY 166 GPAGSP-----GFOGLPGAPGCEAGKPGEQVGGDLGAPGSPGAGEPGTGLPGPP 219
Db 246 GTAGLPGMKHGRGFSGLDAGKADGAPGPKGEPGSPGNGAPGQMGPRGLPGERGRGPP 305
QY 220 GERGGSGRGGFADGVAAGPKGAPAGERSGPGAPGKSPGEAGRPGAGLPGAKGLTSP 279
Db 306 GT---AGARGNDGAVGAAGPPGPT-----GPTGPPGPPGAVGAKGEAGPQCAR----- 350
QY 280 GSPGPDGKTGPPGAGODGRGPPGPGARGQAGVMGFPKGAAGEPGKAGRGVDPGP 339
Db 351 CSEGGQGVRRGPPGPPGAGAGPAGNPGADGQ-----PGAKGANGAPGIA---GAPGFP 401
QY 340 GAVGPAGKDGAGAGQGGPPGAPGAGERGEGGAGSPGFGGLPGSPAGPPGAGKPGEGQVP 399
Db 402 GARGP-----SGPQPSGPPGKGNSE-----PGAPGNKGDGTGAKGPGATGVQGP 449
QY 400 GDLGAPGSPGAPGFPPTGLPGPGERGGPGSRGFPDAGVAGPKGAPAGERGSPGAPGK 459
Db 450 GPAGEEGRKARGGFPGLPGPGERGGPGSRGFPDAGVAGPKGAPAGERGSPGAPGK 509
QY 460 GSPGAGRPAGEAGLPGAKGLTSGSPGPGDGTGPPGAGDGRGPPGPPGARGGAGVM 519
Db 510 GSPGAGEGPGAGLPGAKGLTSGSPGPGDGTGPPGAGDGRGPPGPPGARGGAGVM 569
QY 520 GPPGPKGAAGPPGKAGRGVPGPPGAVGAPGAKDGEAGAGQGGPPGAPGAGERGQGPAGSP 579
Db 570 GPPGPKTAGPPGKAGRGVPGPPGAVGAPGAKDGEAGAGQGGPPGAPGAGERGQGPAGSP 629
QY 580 GFQGLPGAPGPPGAGKPGEGVPGDLGAPGSPGAGE-----PGP----- 620
Db 630 GFQGLPGAPGPPGAGKPGEGVPGDLGAPGSPGAGE-----PGP----- 620
QY 621 -----TGLPDPGPPGSGRGGFPFCADGAVGKPGAPAGERGSPGAPGKSPGGE- 668
Db 690 GAPNDGAKGDTGAPGAPGSGAGFGLQMPGERGAAGLPGKDRGADGPKGADGSPGKD 749
QY 669 -----AGPGEAGLPGAKGLTSGSPGPGDGTGPPGAPGQDGRGPPGPPGARGQA 720
Db 750 GARGLTGPIGPPGAPGAPGKAGPSGPPGPTGARGAPGDRGAPGPPGPPGAPGAPGAP 809
QY 721 GVMGFPKGAAGEFGKAGRGVPGPPGAVGAPGAKDGEAGAGQGGPPGAPGAGERGE--- 776
Db 810 GAPGAPGVGAPGKNGDRGTGAPGAPGIPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 869
QY 777 -----QGPAGSPGQGLPGA-----GPPGEAGKPGE---QGVPGDLGAPGP- 815
Db 870 GHRGFSGLQGPSPGSPGEGSGAGSPGPRGPPGSGAPGKGLNGLPGPIGPPGPR 929
QY 816 -----SGPAG 820
Db 930 GRTGDSGPAG 939
```